Learning and Reversal in the Sub-cortical Limbic System: A Computational Model

A Thesis Presented to the Faculty of Engineering

by

Adedoyin Maria Thompson

in partial fulfillment for the

degree of Doctor of Philosophy

Department of Electronics and Electrical Engineering

University of Glasgow

September 2009

To God, for loving us first.

Mum and Dad, for all your love and support.

To Craig, the one I love.

Acknowledgements

I wish to acknowledge some of the people who have assisted me in this work and made my experience at Glasgow so enjoyable. First and foremost, I would like to thank my supervisor Bernd Porr for believing in me, as well as the guidance he has given throughout the years. His positive energy for science fuelled and motivated me when I was exhausted from hitting walls. I would also like to thank John Williamson, who always invested time and effort in me, for his assistance during my undergraduate and postgraduate years and for the many times he fed and watered me. I am grateful to Professor John O'Reilly, for his subtle ways of telling me to get on with it. I would also like to thank Professor Florentin Wörgötter and Christoph Kolodziejski, who introduced me to the simulator that was used in this work.

A huge thank you goes to my office mates Sven Soell (The German), Rodrigo Garcia (The Mexican), Lynsey McCabe (The Wee Percussionist), Laura Nicolson (The Scottish-Malawian), Colin Waddell (who waddled with twaddle) and Paolo Di Prodi (The Businessman). They all offered much relief from the tedium of research work through long discussions about life and gossip (Bob the builder). To all my friends and the staff at Glasgow University.

I would like to thank Craig Stevenson who always loved and cared for me when all I could think of was science. To Craig's family for making me feel at home especially when I missed home. To my family, for all your support. I am indebted to them for the sacrifices that they made to put me through education.

Thank you.

Abstract

The basal ganglia are a group of nuclei that signal to and from the cerebral cortex. They play an important role in cognition and in the initiation and regulation of normal motor activity. A range of characteristic motor diseases such as Parkinson's and Huntington's have been associated with the degeneration and lesioning of the dopaminergic neurons that target these regions. The study of dopaminergic activity has numerous benefits from understanding how and what effects neurodegenerative diseases have on behavior to determining how the brain responds and adapts to rewards. The study is also useful in understanding what motivates agents to select actions and do the things that they do.

The striatum is a major input structure of the basal ganglia and is a target structure of dopaminergic neurons which originate from the mid brain. These dopaminergic neurons release dopamine which is known to exert modulatory influences on the striatal projections. Action selection and control are involved in the dorsal regions of the striatum while the dopaminergic projections to the ventral striatum are involved in reward based learning and motivation.

There are many computational models of the dorsolateral striatum and the basal ganglia nuclei which have been proposed as neural substrates for prediction, control and action selection. However, there are relatively few models which aim to describe the role of the ventral striatal nucleus accumbens and its core and shell sub divisions in motivation and reward related learning. This thesis presents a systems level computational model of the sub-cortical nuclei of the limbic system which focusses in particular, on the nucleus accumbens shell and core circuitry.

It is proposed that the nucleus accumbens core plays a role in enabling reward driven motor behaviour by acquiring stimulus-response associations which are used to invigorate responding. The nucleus accumbens shell mediates the facilitation of highly rewarding behaviours as well as behavioural switching.

In this model, learning is achieved by implementing isotropic sequence order learning and a third factor (ISO-3) that triggers learning at relevant moments. This third factor is modelled by phasic dopaminergic activity which enables long term potentiation to occur during the acquisition of stimulus-reward associations. When a stimulus no longer predicts reward, tonic dopaminergic activity is generated. This enables long term depression. Weak depression has been simulated in the core so that stimulus-response associations which are used to enable instrumental response are not rapidly abolished. However, comparatively strong depression is implemented in the shell so that information about the reward is quickly updated. The shell influences the facilitation of highly rewarding behaviours enabled by the core through a shell-ventral pallido-medio dorsal pathway. This pathway functions as a feed-forward switching mechanism and enables behavioural flexibility.

The model presented here, is capable of acquiring associations between stimuli and rewards and simulating reversal learning. In contrast to earlier work, the reversal is modelled by the attenuation of the previously learned behaviour. This allows for the reinstatement of behaviour to recur quickly as observed in animals. The model will be tested in both open- and closed-loop experiments and compared against animal experiments.

Declaration

Publications based upon the work contained in this thesis:

Thompson, A. M., Porr, B. and Wo̊rgötter, F. "Learning and Reversal in the Sub-cortical Limbic System: A Computational Model" Adaptive behavior (accepted for publication)

Thompson, A. M., Porr, B. and Wo̊rgötter, F. "Behavioral inhibition during reversal learning in the limbic system: a computational model" Proceedings of the Eighteenth Annual Computational Neuroscience Meeting: CNS 2009, Berlin

Thompson, A. M., Porr, B. Kolodziejski, C. and Wo̊rgötter, F. "Second Order

Conditioning in the Sub-cortical Nuclei of the Limbic System" Proceedings of the LNCS/CNAI: SAB 2008, Japan

Thompson, A. M., Porr, B., Egerton. A and Wo̎rgötter, F. "How bursting and tonic dopaminergic activity generates LTP and LTD" Proceedings of the Seventeenth Annual Computational Neuroscience Meeting: CNS 2007, Toronto

Thompson, A. M., Porr, B. and Wo̎rgötter, F. "Stabilising Hebbian learning with a third factor in a food retrieval task" Proceedings of the LNCS/CNAI: SAB 2006, Rome

Contents

List of Acronyms

BLA ..... basolateral nucleus of the amygdala

CNA ..... central nucleus of the amygdala

CR ..... conditioned response

CS ..... conditioned stimulus

CS+ ..... CS paired with an unconditioned stimulus

CS- ..... CS not paired with an unconditioned stimulus

D1 ..... type 1 dopamine receptor

D2 ..... type 2 dopamine receptor

D3 ..... type 3 dopamine receptor

D4 ..... type 4 dopamine receptor

D5 ..... type 5 dopamine receptor

DA ..... dopamine

DAergic ..... dopaminergic

DS ..... discriminative stimulus

EC ..... enthorhinal cortex

ECB ..... endocannabinoid

EPSP ..... excitatory post synaptic potential

fLTD ..... full-LTD

GABA ..... gamma-aminobutyric acid

Glu ..... glutamate

GP ..... globus pallidus

GPe ..... globus pallidus external

GPi ..... globus pallidus internal

HFS ..... high frequency stimulation

HPC ..... hippocampus

ISI ..... interstimulus interval

LH ..... lateral hypothalamus

LI ..... latent inhibition

LMS ..... least mean square

LTD ..... long term depression

LTP ..... long term potentiation

M1 ..... muscarinic M1 receptor

MD ..... medio dorsal nucleus of the thalamus

mGlu ..... metabotrophic type glutamatergic

MSNs ..... medium spiny neurons

NAc ..... nucleus accumbens

NMDA ..... N-methyl-D-aspartic acid

NO ..... nitric oxide

PCA ..... Pavlovian conditioned approach

PFC ..... prefrontal cortex

PIT ..... Pavlovian instrumental transfer

pLTD ..... partial-LTD

pLTD-MD .... partial-LTD-MD-feedforward

PPTN ..... pedunculopontine tegmental nucleus

PVLV ..... primary value learned value

RL ..... Reinforcement learning

SLG ..... Schmajuk, Lam and Gray

SNc ..... substantia nigra pars compacta

SNr ..... substantia nigra pars reticulata

STDP ..... spike timing dependent plasticity

STN ..... subthalamic nucleus

TD ..... temporal difference

UR ..... unconditioned response

US ..... unconditioned stimulus

VA ..... ventral anterior

VP ..... ventral pallidum

VTA ..... ventral tegmental area

List of Figures

List of Tables

Chapter 1

Introduction

The ability of an agent to adapt according to changing conditions in the environment is a necessity for survival. For example, a squirrel finds nuts under a tree. It learns to associate the tree with nuts and always goes to the tree when it searches for nuts. At some point, there are no more nuts under the tree. How does the squirrel stop going to that tree to look for nuts whilst still maintaining an association between the tree and the nuts so that in the future, when the tree starts reproducing nuts, the squirrel may return and find nuts under the tree again? This is an example of reversal learning. When a stimulus-reward (seeing the tree - getting nuts) contingency changes, an agent's behaviour towards the stimulus which once predicted the reward changes accordingly. Biological agents can demonstrate such behavioural flexibility by inhibiting appetitive behaviour when the incentive value of the conditioned stimulus (CS) that predicts the reward changes.

One popular model used in prediction and control is the actor-critic model (Sutton and Barto, 1998). In this model, the critic uses a learning rule, usually a temporal difference (TD) method, (Sutton and Barto, 1982, 1987, 1990) to calculate the error signal which is then used to train the actor. The TD error becomes positive when an unexpected reward is obtained. During reversal, when the reward is omitted, a negative error is produced which depletes the learned actions. Such depletion of learned actions do not ac-

count for animal behaviour such as rapid reacquisition (Pavlov, 1927; Napier et al., 1992) and led to suggestions that learned associations are not simply eliminated (Rescorla, 2001) during omission. The process of "*learning*" then "*unlearning*" then "*learning*" stimulus - response associations is an inefficient and biologically unrealistic mechanism that is currently implemented in many computational models. This conceptual framework has also been reviewed by both Bouton (2002) and Rescorla (2001) who argue against "unlearning" during extinction. A more efficient way would be to suppress or disable the actions so that they can be quickly reactivated when necessary.

This thesis presents a biologically motivated computational model of the sub-cortical nuclei of the ventral striatal circuitry at a systems level. The ventral striatum comprising the nucleus accumbens, plays a role in processing rewards and has been labelled as a "*limbic-motor*" interface. It is the region whereby "*motivational-emotional determinants of behaviour become transformed into actions*" (Mogenson et al., 1980). The model proposes that the ventral striatum and surrounding circuitry provide a mechanism of enabling and disabling learned action systems as required when rewards are presented and omitted respectively. It will be shown how the model learns an association between a stimulus and the action system that results in a reward. When the reward is omitted, the model makes adjustments accordingly, not by the more popular method of eliminating learned associations (O'Reilly et al., 2007), but by using a feed-forward switching mechanism to disable the action system.

The next section introduces the concept of control in embedded agents, the most common of which are animals embedded in their environments. Animal learning is discussed and classified as open- and closed-loop systems. The history of adaptive systems is long, therefore, a few adaptive elements which are associated with animal learning will be introduced and will be linked to actor-critic models and biological models of action selection and goal directed behaviours.

1.1 Adaptive Control and Animal Learning

A control system comprises a network of components which form a system configuration that produces a desired response (Dorf and Bishop, 2005). A controller can be used in both open- and closed-loop systems. The behaviour of the controlled system is usually affected by disturbances. For a closed-loop control system, feed-back provides information about such disturbances to the controller. In open-loop systems, a controller generates an output using inputs that are not influenced by its output i.e. a system that controls a process without using feed-back.

In adaptive control, there are a number of changes that occur which often lead to changes in the parameter values that affect the performance and stability of controlled systems. Adaptive controllers have been used to make adjustments when such parameters change. Adaptive networks became popular because they were capable of exhibiting a variety of behaviours observed in animal learning (Rescorla, 1972; Sutton and Barto, 1990; Klopf et al., 1993). These have been used as computational analogs for animal learning.

Animal learning implements methods by which animals predict and respond to important events in the environment. It has been classified according to two experimental procedures known as classical or Pavlovian conditioning and instrumental or operant learning. In chapter 2, a variety of experimental studies involved in animal learning and behaviour will be collated and used to develop a computational neural network model which aims to describe certain processes involved in animal learning. In the following sections, the history of classical and instrumental conditioning are summarised and classified in terms of open- and closed-loop systems.

1.1.1 Classical Conditioning as an Open-loop System

During the end of the early twentieth century, the Russian physiologist, Ivan Petrovich Pavlov while investigating the digestive function of dogs, noticed

that dogs salivated to the sound of a bell presented before they received food. This prompted him to produce his definition of the basic laws that govern the creation and extinction of the responses he termed "*conditional reflexes*" (Pavlov, 1927). His work has led to the well known concept of Pavlovian or classical conditioning. Classical conditioning is an elementary form of associative learning in which a previously neutral stimulus, commonly referred to as a conditioned stimulus (CS) which occurs prior to a primary reward or unconditioned stimulus (US), generates a conditioned behaviour or conditioned response (CR) similar to the behaviour exhibited when the US occurs (Pavlov, 1927). This response generated in event of the US is known as the unconditioned response (UR). Stimulus substitution occurs when the CR is elicited before the US is presented and therefore before the UR. The CR demonstrates that the animal has learned to predict the delivery of the US.

The classical conditioning paradigm is an open-loop procedure because the stimuli delivered are not contingent on the animal's behaviour. The classical conditioning paradigm as an open-loop procedure in an embedded agent is illustrated in Fig. 1.1A. An agent receives input signals which correspond to the CS and the US. Its output is represented by the unconditioned response which slowly becomes replaced by the conditioned response. The agent obtains a US or reward which is not dependent on the responses it makes. Chapter 3 presents a variety of open-loop phenomena in which the computational model that was developed in chapter 2 is tested. The model is capable of demonstrating a range of classical conditioning phenomena.

1.1.2 Instrumental Conditioning as a Closed-loop System

The difference between instrumental and classical conditioning was identified by Skinner, whose ideas that operant behaviours in animals are instrumental and bring about consequences, are based on Thorndike's law of effect (Gross, 2001). Instrumental conditioning is regarded as a closed-loop procedure be-

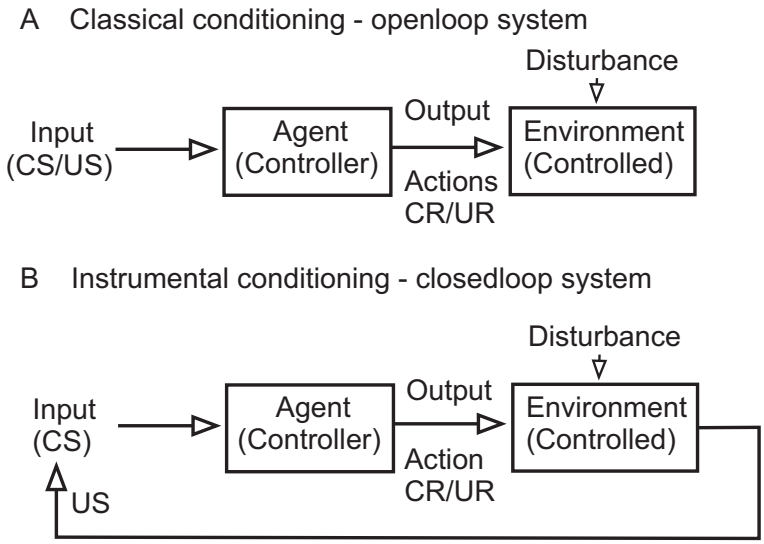

Figure 1.1: A) Classical conditioning as an open-loop system. B) Instrumental conditioning as a closed-loop system. (Dorf and Bishop, 2005).

cause the stimuli obtained by the animal depend on the its actions (Klopf et al., 1993). This is illustrated in Fig. 1.1B. The feed-back indicates that the US or reward obtained by the agent is dependent on the response it elicits. In chapter 4, the computational model is tested in a variety of closed-loop behavioural experiments. Its performances in reward seeking tasks are compared to animal experiments. The results generated by the model have similarities with the results presented from the animal experiments.

Edward Thorndike built puzzle boxes in which cats were placed and had to learn to operate a lever to exit the box (Thorndike, 1911; Gross, 2001). By doing so, they obtained a reward located outside the box but which had been visible from inside the box. Although the cats initially struggled, with repetition, they required less time to make exits and get the reward. Thorndike proposed that animals learn from "*trial and error*" and associations between the stimulus and response are "*strengthened*" by the reward and "*weakened*" otherwise. Reinforcers and punishers were defined as the consequences that "*strengthen*" or "*weaken*" behaviours respectively. Although Bandura (1977) described reinforcers as a principally informative and motivational operation

rather than as a physical response strengthener, Reinforcement learning is based on Thorndike's ideas of law of effect (Gross, 2001).

1.1.3 Reinforcement Learning

Reinforcement learning (RL) originates from trial and error learning, temporal difference methods and aspects involving optimal control (Sutton and Barto, 1998). Reinforcement learning, as defined by Sutton and Barto (1998), is learning how to maximize a numerical reward by mapping situations to actions. During RL, an embedded agent performs actions in its environment. The RL agent is not told which actions lead to a maximum reward but through trial and error, it learns to select these actions based on its previous experiences (Sutton and Barto, 1998; Porr and Wörgötter, 2005), using evaluative feed-back which indicates the "*goodness*" of the action taken. Initially implemented for goal directed learning and decision making, it has been classified according to its ability to solve Reinforcement learning problems (Sutton and Barto, 1998). Reinforcement learning can be a useful tool in solving engineering control problems (Houk et al., 1995; Barto et al., 1990). It will be shown how certain mechanisms used by Reinforcement learning methods are implemented by the computational model developed in this work. In addition however, the model is modified so that acquired associations are not unlearned and a feed-forward mechanism is used which facilitates and attenuates the action system.

Numerous attempts have been made to reproduce the effects of classical and instrumental conditioning in real-time computational models (Sutton and Barto, 1990; Klopf et al., 1993; Houk et al., 1995; Balkenius and Morén, 1998; Suri and Schultz, 1999). Some of the rules are addressed which show how adaptive elements have been linked to animal learning.

1.2 Neuronal Analogs of Animal Learning and Behavior

In this section a few rules and models which have made rather influential contributions to the study of animal learning and adaptive control will be discussed. These rules include Hebbian learning, the Rescorla-Wagner model, the Sutton-Barto model and the temporal difference model. The goal is to illustrate how each rule is related to the next and this in turn will demonstrate briefly how neural models of animal learning and control may have evolved. The first rule that will be addressed was inspired from biology and has been used in adapted versions to explain how associations between neurons are formed and therefore how learning can be achieved.

1.2.1 Hebbian Learning

Hebbian learning is a very popular correlation based learning rule which originated from Donald Hebb's postulate. It states:

> Any two cells or system of cells that are repeatedly active at the same time will tend to become "associated", so that activity in one facilitates activity in the other. (Hebb, 1949) (p.70).

In other words when pre- and postsynaptic neurons both undergo activity, the connectivity between them becomes strengthened.

A neuronal unit that implements the Hebbian rule is shown in Fig 1.2A. There are n input pathways or presynaptic terminals indexed by i where $i = 1, ..., n$ and output signal y(t) which represents the postsynaptic neuron. Each input pathway is associated with a weight w_i which changes according to a mathematical representation of Hebb's postulate.

$$\Delta\omega_i(t) = cx_i(t)y(t) \tag{1.1}$$

Where c is a learning rate constant that determines the rate by which ω_i changes. The Hebbian rule accounts sufficiently for the stimulus substitution view of classical conditioning (Sutton and Barto, 1981) but does not specify the importance of timing between the pre- and postsynaptic events. However popular this postulate was, it was not sufficient in itself to account for the temporal aspects involved in classical conditioning. This led to modified versions of the rule including the differential Hebbian learning rule. The differential Hebbian learning rule has been integrated in the model presented in this thesis. It will be addressed in chapter 3 during the development of the computational model. The following section introduces the Widrow-Hoff rule which has some similarities with the Hebbian learning rule.

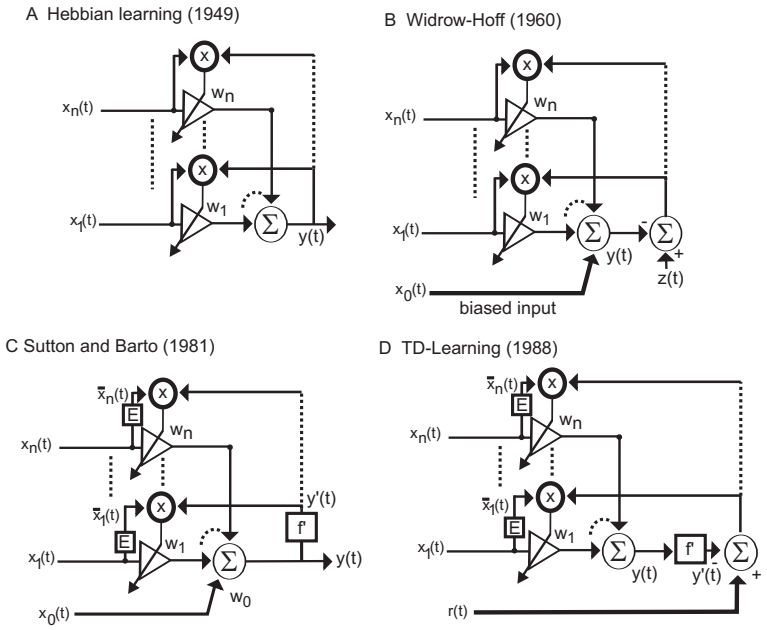

Figure 1.2: A) Hebbian learning, B) Widrow-Hoff rule, C) Sutton and Barto (Sutton and Barto, 1981) and D) TD-learning (Sutton, 1988). f' represents the function which generates the difference of the output y over a duration between t and t-1. Adapted from Porr and Wörgötter (2005); Kolodziejski et al. (2008).

1.2.2 The Widrow-Hoff Rule

The Widrow-Hoff rule, also known as the delta rule, is an elegantly simple rule that was presented in an adaptive element by Widrow and Hoff (1960). It has been implemented in a variety of practical applications including antennas, telecommunications and filters. In signal processing, it is referred to as the least mean squares (LMS) algorithm.

The adaptive neuron model (Fig 1.2B) that implements this rule, comprises a set of input signals (x_i) indexed by $i = 1, ..., n$. An additional input signals (x_0), makes up the bias input and is set to a value that represents the threshold value of the neuron. The other inputs are weighted and these weights change according to:

$$\Delta\omega_i(t) = c[z(t) - y(t)]x_i(t) \tag{1.2}$$

Here, a specialized teacher or "boss" ($z(t)$) signals the desired output. Both $z(t)$ and $x_i(t)$ are real numbers. The output ($y(t)$) represented by the sum of the weighted inputs as $y(t) = \sum_{i=1}^{n} x_i w_i$, is used to calculate the error signal which is formed from the difference between the desired and actual output signals. This can be associated with the Hebbian learning rule wherby the postsynaptic activity is replaced instead by the error signal $[z(t) - y(t)]$. The weights converge such that the output tends towards the desired value and the error tends towards zero.

An error surface is generated as a function of the weighted inputs. A gradient descent method is used to identify which direction the weights can be adjusted so that the gradient of the error surface is at its minimum. However, these local minima do not guarantee that the overall error surface gradient generates a global minimum. Widrow and Hoff (1960) show that the partial derivatives of the error sequence with respect to the weights is proportional to the error signal (Widrow and Hoff, 1960).

The Widrow and Hoff rule is a form of supervised learning because it requires

input from an external supervisor. Coincidentally, the Widrow and Hoff rule which was developed for engineering solutions follows a similar format to the Rescorla-Wagner model which was implemented to account for paradigms of classical conditioning (Sutton and Barto, 1981). The Rescorla-Wagner rule is described next.

1.2.3 The Rescorla-Wagner Model

Rescorla-Wagner's model of classical conditioning was a very influential model that was designed to account for certain features of classical conditioning. The model has an error correction element and is similar to the learning algorithm of Widrow and Hoff (1960) (Sutton and Barto, 1987). It attempts to predict how the "associative strength" between the CS and US changes over a number of trials according to the occurrences of the stimuli and to the unpredictability of their occurrence. The Rescorla-Wagner model is therefore a trial level model and the predictions made by the model are not dependent on the temporal relationships between the CS and US events. The associative strengths change based on the differences between the actual and expected US. The actual US level is represented by λ, and the expected or predicted US is obtained from the sum of all the associative strengths of the CS. On every trial, the presence or absence of the ith CS is indicated by $CS_i = 1$ and $CS_i = 0$ respectively. The associative strength of each CS_i combined with the CS_i can be represented by w_{CS_i}. The sum of all the associative strengths becomes $\bar{w} = \sum_i w_{CS_i}$. In the Rescorla-Wagner rule, the associative strength for every present CS_i changes according to

$$\Delta w_{CS_i} = c[\lambda - \bar{w}]CS_i \tag{1.3}$$

This rule, although constructed for a very different purpose, takes a similar form to the Widrow-Hoff rule of Eq. 1.2 (Sutton and Barto, 1981). In the Rescorla-Wagner rule, the change in the associative strength occurs with respect to the trial number rather than the time. There are, however, short-

comings of the Rescorla-Wagner rule. As a trial-level model, the Rescorla-Wagner model is incapable of making predictions about the intratrial temporal relationship effects on learning (Sutton and Barto, 1990). In addition, the Rescorla-Wagner model needs to be modified so as to make predictions about another aspect of classical conditioning known as second-order conditioning.

Learning in real-time mechanisms is dependent on the occurrence as well as the timing of significant events such that the associative strengths can change on a moment by moment basis within trials. The Rescorla-Wagner and Widrow-Hoff rules do not implement real time mechanisms because learning is only determined by the order of the input, output and learning signals rather than their associations in time.

1.2.4 The Sutton-Barto (SB) Model

Sutton and Barto (1981) developed an adaptive element analog of classical conditioning that is a real-time extension of the Rescorla-Wagner model. It accounted more closely for animal learning theories than corresponding adaptive networks studied at the time. In addition to its ability to effectively predict reinforcement, the model is also capable of solving stability issues (Sutton and Barto, 1990).

Fig 1.2C illustrates a representation of the neuron like unit that utilises the SB rule. This unit has certain differences to the neuronal representation of the Hebbian model. There are n input pathways for inputs x_i where $i = 1, ..., n$. There is also an x_0 input which represents the US input and which has a fixed positive weight w_0 in its pathway. The $x_1, ..., x_n$ inputs represent the CS inputs whose weights are adaptable. The input signals generate a stimulus trace so as to define periods of eligibility. When a stimulus occurs at a certain time $x_i(t)$, the stimulus generates a prolonged trace for a duration after t. This prolonged stimulus trace is represented by $\bar{x}_i$ and is obtained from the weighted average values of x_i. This is represented in Fig. 1.2C by passing the input signals x_i through the respective function labelled E.

$$\bar{x_i}(t) = \beta \bar{x_i}(t-1) + (1-\beta)x_i(t-1) \tag{1.4}$$

where $0 < \beta < 1$. y(t), the output at time t is the sum of all the weighted inputs. The weights change in the SB-model according to:

$$\Delta w_i(t) = c\bar{x_i}(t)(y(t) - y(t-1)) \tag{1.5}$$

$y(t) - y(t-1)$ is the difference between the current and previous output value and is generated by the funtion f' to represent $y'(t)$ in Fig. 1.2C.

Any active input x_i generates a change in the output pathway and its respective weight becomes eligible to change over the duration of the input stimulus trace. Eligible weights change depending on the discrete rate of change of the output y(t). The SB rule can be identified as a version of the Hebbian learning rule whereby the weight modification depending on the input and output signals, corresponds to the change that occurs due to a correlation between the input trace and the change in the output signal (Sutton and Barto, 1981).

The Sutton-Barto model accounts for a variety of classical conditioning effects including blocking, conditioned inhibition and certain effects of intratrial temporal relationships. However, the SB model predicts weak or inhibitory conditioning for short interstimulus intervals (ISI) (Sutton and Barto, 1990). This in turn led to the development of the temporal-difference (TD) model. Another model that addressed this issue was the correlation based differential Hebbian learning algorithm known as Isotropic sequence order (ISO) learning (Porr, 2004). The TD method is discussed next.

1.2.5 The Temporal-Difference Model

The TD model is introduced according to its definition by Sutton and Barto (1998). It is then re-represented as a neuronal analog similar to that pre-

sented in Porr and Wörgötter (2005). TD methods are defined in terms of the return, the value function and the policy (Sutton and Barto, 1998). The return (R_t) is regarded as the discounted sum of rewards (r) received after every time step (t) and is represented as follows.

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + ... \tag{1.6}$$

where $0 \leq \gamma \leq 1$, and γ is the discount factor which places a higher value on immediate rewards than the value placed on future rewards. The return can be simplified to give

$$R_t = r_{t+1} + \gamma R_{t+1}. \tag{1.7}$$

The policy (π) is the mapping of states and actions to the probability of taking an action a when in the state s (Sutton and Barto, 1998). The value of a state ($V^\pi(s)$) is the expected return when starting in that state and following a particular policy π (Sutton and Barto, 1998). It is defined using $E_\pi\{\}$ to represent the expected value as follows

$$V^\pi(s_t) = E_\pi\{R_t | s_t = s\}. \tag{1.8}$$

The value of the state is re-represented using Eq. 1.7 to give

$$V^\pi(s_t) = E_\pi\{r_{t+1} + \gamma R_{t+1} | s_t = s\}. \tag{1.9}$$

This is be approximated using the estimate of the immediate next state as follows

$$V^\pi(s_t) = E_\pi\{r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s\}. \tag{1.10}$$

The value of the state is updated incrementally by adding it to the difference between the actual reward R_t and current value of the state $V(s_t)$

$$V(s_t) \leftarrow V(s_t) + \alpha \left[R_t - V(s_t)\right]. \tag{1.11}$$

α determines the step size by which the value function is updated. If the actual return R_t is unknown, the update is obtained with the assumption

that the value of the next state Vs_{t+1} is an accurate estimate of the expected return downstream to this state.

$$V(s_t) \leftarrow V(s_t) + \alpha \left[r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \right]. \tag{1.12}$$

The value of the state is represented in its neuronal analog as a function of time (t) to give the output $y(t)$. The neuronal representation of the temporal-difference (TD) model is shown in Fig. 1.2D. It is an extension of the SB model and was designed to deal with the problems that the SB model faced (Sutton and Barto, 1990). The $x_1, ..., x_n$ inputs represent the CS inputs whose weights change in the TD-model according to the following rule

$$\Delta w_i(t) = c[w_0(t)x_0(t) + \gamma y(t) - y(t-1)]\bar{x}_i(t). \tag{1.13}$$

In this case y(t) represents only the sum of the weighted CS inputs $y(t) = \sum_{i=1}^{n} w_i(t)x_i(t)$ and so the output is not driven directly by the US pathway. The US or $w_0(t)x_0(t)$ pathway can be referred to as the reward or reinforcement signal r. It is combined with the difference between the current and previous output $y(t) - y(t-1)$ and the TD-error signal, $\delta(t)$ is represented in its more recognisable form

$$\delta(t) = r(t) + \gamma y(t) - y(t-1). \tag{1.14}$$

Like the SB-rule, the weight change is also dependent on the difference between the current and previous output signal. However, while the output signal is adapted so that it is not directly influenced by the $w_0(t)x_0(t)$ pathway, the weight change on the other hand is dependent on this US signal which forms an element of the reinforcement.

1.2.6 Actor-Critic Models

Actor-critic methods have been implemented in control theory as an approach to solving nonlinear control problems (Houk et al., 1995; Barto, 1995; Witten, 1977; Sutton, 1984). They have become influential in animal learning concepts (Porr and Wörgötter, 2005), and have been implemented as neuronal models for prediction and control (Houk et al., 1995). They comprise two separate units called the actor and the critic. As the name suggests, the actor selects actions according to the policy which defines the agent's behaviour while the critic is the estimated value function which can be represented by any learning algorithm but usually takes the form of the TD-error signal Barto (1995); Suri and Schultz (1999); Porr and Wörgötter (2005). It evaluates and criticizes the actions made by the actor and in addition uses this signal to train itself. The actor-critic architecture as an adapted control system is presented in the form of a block diagram in Fig 1.3.

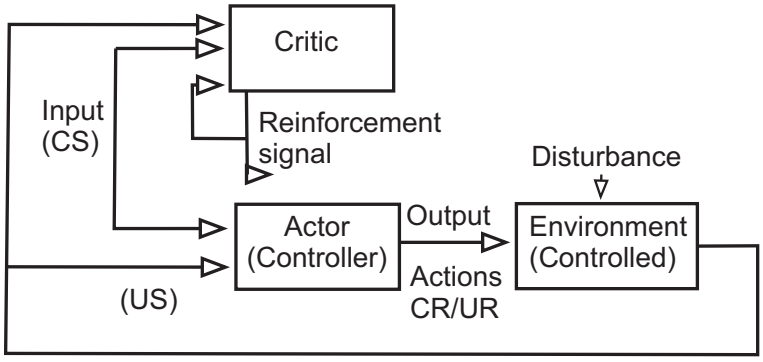

Figure 1.3: A block diagram representation of the actor-critic architecture. Modified from (Barto, 1995; Porr and Wörgötter, 2005).

By implementing reinforcement learning, one advantage of the actor-critic model is that actions can be selected with minimal computational requirements (by using algorithms such as the TD-error signal). In addition, they are capable of acquiring the optimal probabilities for selecting various actions. Such actor-critic architectures rely on the *return maximization principle* whereby the best actions are chosen so that the maximum expected

return is obtained. When actions are made, evaluative feed-back is obtained from environment which returns a number that ranges from negative to positive values. The actor makes decisions by making a comparison against all possible actions. The weight for the selected action changes according to the Effective Reinforcement signal.

The Adaptive-Critic

The adaptive critic was developed by Sutton (1988) and implemented by Barto et al. (1983) as a neuronlike adaptive element. It learns to predict reinforcing events (Houk et al., 1995). It was identified as an approach to solving the temporal credit assignment problem (Barto, 1995; Houk et al., 1995). In animal learning, the credit assignment problem is the problem of allocating reinforcement to the correct synapse and the temporal credit assignment problem is the problem of allocating credit to the right synapse at the right time. The Effective Reinforcement signal generated by the adaptive critic is the same as the TD-error $\delta = r(t) + \gamma y(t) - y(t-1)$.

Actor-critic architectures which have been implemented as neuronal models for prediction and control generally represent the connectivity between the basal ganglia and the cortex. There are many computational models that have been developed which suggest how the cortex and basal ganglia play a role in prediction, control, action selection , decision making and goal directed behaviours (Brown et al., 1999; Houk et al., 1995; Gillies and Arbuthnott, 2000; Gurney et al., 2001a; Joel et al., 2002; Prescott et al., 2006; Porr and Wörgötter, 2005). In the following section, the basal ganglia according to the actor-critic architecture is introduced briefly. In the following chapter, the biological setting including the basal ganglia are discussed.

1.2.7 Actor-Critic Models and the Basal Ganglia

The basal ganglia are a group of nuclei located at the base of the forebrain comprising parallel loops and structurally and functionally distinct circuits

which connect to the cortex. These create a large sub-cortical network which is involved in both voluntary and higher brain functions (Alexander and Crutcher, 1990; Nambu, 2009). The basal ganglia have traditionally been viewed as a motor control system. The evidence for this is that lesion and damage to this region almost always results in movement disorders (Clark and Boutros, 1999).

Certain regions of the basal ganglia receive projection from dopaminergic neurons which release the neuromodulator dopamine. These dopaminergic neurons have shown very interesting firing patterns in relation to rewards and reward predicting stimuli. There are a variety of effects for which the study of the role of dopaminergic activity in the basal ganglia is essential. One such example can be observed Parkinsons disease, a hypokinetic disorder which results from degeneration of dopaminergic neurons. This work focusses on the role of dopaminergic activity as a mechanism that influences plasticity in the limbic system. Dopaminergic neurons are activated by unpredictable rewards. This activity slowly decreases as the reward becomes predictable. They are also activated by reward predicting stimuli and show a depression when rewards are omitted (Ljungberg et al., 1992; Schultz et al., 1997; Suri and Schultz, 1999). These DA firing patterns have been associated with the TD-error implemented in RL (Sutton, 1988; Montague et al., 1996; Schultz et al., 1997; Suri and Schultz, 1999; Joel et al., 2002). The TD-error has been used by the adaptive-critic in RL actor-critic models.

While the studies conducted by Schultz and Dickinson (2000) initiated the development of Reinforcement learning models of the basal ganglia, Houk et al. (1995) were among the first to match the actor-critic architecture to the basal ganglia and surrounding circuits (Joel et al., 2002; Porr and Wörgötter, 2005). Houk et al. (1995) proposed that two different striatal modules adopt different roles depending on their characteristic afferent projections. While the striosomal modules which include connections from the striatal spiny neurons to DA neurons function as the adaptive critic, the matrix modules assume the role of the actor.

A variety of other actor-critic models exist including the model by Suri and Schultz (1999), in which the TD algorithm was adapted to accommodate a timing mechanism that effectively reproduced the timed depression of DA activity (Joel et al., 2002).

The work by Houk et al. (1995) and Suri and Schultz (1999) are two examples of attempts that aim to describe how actor-critic architectures can be used to explain how the basal ganglia performs prediction and control. These two models are among a variety of Reinforcement learning actor-critic methods that interpret DA activity as a temporal difference (TD) error (Sutton and Barto, 1982, 1987, 1990; Joel et al., 2002). The TD-error signal generated by the critic represents the difference between the current and previous estimate of the return. It is used to control the actor so that the stimuli which lead to maximum rewards are utilized. The actor is "*taught*" to learn new sensor motor associations guiding the agent to the reward. When expected rewards are omitted, a negative value is generated so that the previously learned associations are eliminated. The models of the basal ganglia attempt to simulate certain functionalities of the basal ganglia at different levels (Brown et al., 1999; Houk et al., 1995; Gillies and Arbuthnott, 2000; Gurney et al., 2001a; Joel et al., 2002; Prescott et al., 2006; Joel et al., 2002; Porr and Wörgötter, 2005). Many of these models focus on the dorsal striatum and have been associated with prediction, control, action selection , decision making and goal directed behaviours. However, the ventral striatum is a major part of the limbic system and can be identified as the reward system of the brain. It is associated with emotions, motivation, behavioral flexibility and reward based behaviors which in turn are linked to conditions such as drug addiction. Reward based learning in the limbic system is discussed next.

1.3 Reward Based Learning and Unlearning in the Limbic System

The limbic system as the reward system of the brain has been modelled so far as a modified classical TD learner (Schultz, 1998; Dayan, 2001) whereby the circuitry surrounding the core and shell are related to the actor and value systems respectively. The TD-rule became popularly linked to the activity of midbrain dopaminergic neurons and have been implemented as the critic in many actor-critic architectures (Suri and Schultz, 1999; Sutton, 1984). An error signal maps to DA generated by dopaminergic neurons which is released as a global value, deciphers the general direction of plasticity of its target structures including the shell and the core. In this model both the core and shell undergo long term depression (LTD) as soon as the reward has been omitted. As mentioned earlier, this means that when a learned association no longer leads to a reward, the agent unlearns the association. This seems to be an inefficient way of learning and adapting because rewards might recur and the actor must once again "*re-learn*" the associations it previously wiped out. A more efficient way is to suppress the actions so that they can be quickly reactivated when necessary. It is known from animal experiments that learned behaviours can undergo rapid reacquisition as soon as the unconditioned stimulus (US) is reintroduced (Pavlov, 1927; Napier et al., 1992). This suggests that behaviours are suppressed rather than unlearned. The current work proposes an alternative mechanism to unlearning in the limbic system when rewards are omitted.

The TD error has been utilised in numerous computational models (Montague et al., 1996; Dayan and Balleine, 2002; O'Reilly et al., 2007). A positive and negative prediction error are used to encode an increase and pause of DA neuron activity respectively. Cragg (2006) has argued that the error between expected and omitted rewards does not quantitatively correlate with the pause in DA activity. In addition, the low baseline firing rates of DA neurons makes it difficult for recipient units to detect and decode the pause in DA activity during omission (Daw et al., 2002; Cragg, 2006) Rather than

implementing a pause in tonic firing patterns, the current model employs a rise in tonic activity to to encode reward omission. This rise in tonic DA activity acts on D2 receptors and is assumed to generate LTD in the current model. Tonic DA has been implemented in Gurney et al. (2004) as an attenuating mechanism of salient stimuli by activating D2 receptors. These two assumptions can function in synchrony in such a way that the attenuation of salient stimuli by tonic DA activity might result in LTD occuring heterosynaptically at synapses.

This chapter began by introducing adaptive elements as neural representation for animal learning. Animal learning has been classified in terms of Pavlovian and instrumental conditioning which have been used as a basis for understanding how animals learn to predict future rewards. The neuronal analogs that were addressed ended with the actor-critic architecture. The actor-critic method has been proposed as a neuronal representation of the basal ganglia in prediction, control and action selection (Houk et al., 1995; Suri and Schultz, 1998, 1999; Brown et al., 1999; Joel et al., 2002). There are numerous RL computational models of the basal ganglia. However, there are also a range of other non-RL computational models that have been developed which suggest its role in action selection (Redgrave et al., 1999a; Gurney et al., 2001a,b; Prescott et al., 2006) sequence learning (Berns and Sejnowski, 1998) and prediction (Houk et al., 1995; Joel et al., 2002). Comparatively fewer models have been proposed which focus on the role of the limbic-regions in motivation and behavioural flexibility.

This thesis proposes a computational model of the sub-cortical limbic system which in particular, implements circuitry of the ventral striatum in reward based learning and reversal learning. Reward functions and appetitive motivated behaviours have been associated with the mesolimbic dopamine (DA) neurons (Wise et al., 1978; Wise and Rompre, 1989) originating from the ventral tegmental area (VTA) which target the nucleus accumbens (NAc) of the ventral striatum. The current model departs from the standard actor-critic architecture with two major characteristics. The current model maintains learned associations when a stimulus no longer predicts a reward but enables

adjustment in response by employing a feed-forward switch. In chapter 5, the computational model is reintroduced whereby the NAc circuitry is also presented so that the core and shell networks function in accordance with the actor and critic respectively. This is so that the model can be associated with and compared against the standard actor-critic architecture.

The model's performance in an open-loop reacquisition test and closed-loop behavioural reversal learning experiments will be compared against actor-critic versions of the model. The models are categorised according to whether or not a feed-forward switching mechanism is implemented to disable actions, or the unlearning rate of the actor component of the model. It will be shown how the current model, which implements the feed-forward switching mechanism and or a slower rate of unlearning in the actor, performs better than the standard actor-critic equivalent.

This thesis closes with a discussion in which other computational models are compared against the current model and concludes with suggestions for future work. In the next chapter, the biological setting is presented. It commences with a brief introduction to the basal ganglia.

Chapter 2

The Biological Setting

The previous chapter briefly described how a variety of computational models of the basal ganglia propose to play a role in prediction, control, action selection, decision making and goal directed behaviours (Brown et al., 1999; Houk et al., 1995; Gillies and Arbuthnott, 2000; Gurney et al., 2001a; Joel et al., 2002; Prescott et al., 2006). In this chapter, the biological setting is establised and commences with and introduction to the basal ganglia. The striatum and the subthalamic nucleus are the principal input components of the basal ganglia (Clark and Boutros, 1999; Redgrave et al., 1999a). The striatum can be divided into the dorsal and ventral division. The dorsal striatum is involved in action selection and motor functions while the ventral region comprising the nucleus accumbens (NAc), is involved in motivation, reward and attention (Schotanus and Chergui, 2008).

This chapter discusses mechanisms involved in information processing and synaptic plasticity in the ventral striatum and proposes how the nucleus accumbens plays a role as the limbic-motor interface (Mogenson et al., 1980). There are currently a few computational models which provide detailed description on the mechanism by which the nucleus accumbens contributes to the basal ganglia's functionality in action selection and behavioural flexibility. This chapter intends to bring the reader's attention to some of the behaviours and functions that the nucleus accumbens has been implicated

in. These behaviours range from Pavlovian conditioning to attentional set-shifting and behavioural flexibility. The underlying goal is to develop a broad understanding of its overall role so as to produce a computational model that mirrors some of the ventral striatal circuitry at a systems level. First, the basal ganglia and dorsal striatum are described. This will then lead to a detailed discussion of the vertebrate ventral striatum and its relevant connectivity and functionality.

2.1 The Basal Ganglia

A representation of the basal ganglia circuitry is illustrated in Fig. 2.1. The basal ganglia comprise a variety of nuclei including the striatum, the external (GPe) and internal (GPi) segments of the globus pallidus (GP), the substantia nigra pars reticulata (SNr), the subthalamic nucleus (STN) and the dopaminergic neurons of the substantia nigra pars compacta (SNc), and ventral tegmental area (VTA). The output nuclei of the basal ganglia are the SNr and the GPi. The GPi and GPe are homologous to the rat entopenducluar nucleus (EP) and globus pallidus respectively Gurney et al. (2004). The basal ganglia receive excitatory inputs from the cerebral cortex and project integrated responses back to the cerebral cortex (Clark and Boutros, 1999).

The striatum, which is divided into dorsal (caudate nucleus and putamen) and ventral (nucleus accumbens (NAc)) parts, is a major input structure to the basal ganglia. The striatum receives inputs from the cerebral cortex as well as the thalamus, and efferents inhibitory GABAergic, gamma-aminobutyric acid (GABA) fibres on the GP, SNr and ventral pallidum (VP). The GP, SNr and VP project to the mediodorsal (MD) and ventral anterior (VA) nucleus of the thalamus which in turn project back to the frontal cortex. Additional important projections to the striatum include the dopaminergic neurons of the SNc and VTA (Utter and Basso, 2008; Joel et al., 2002; Kandel et al., 1991; Alexander and Crutcher, 1990).

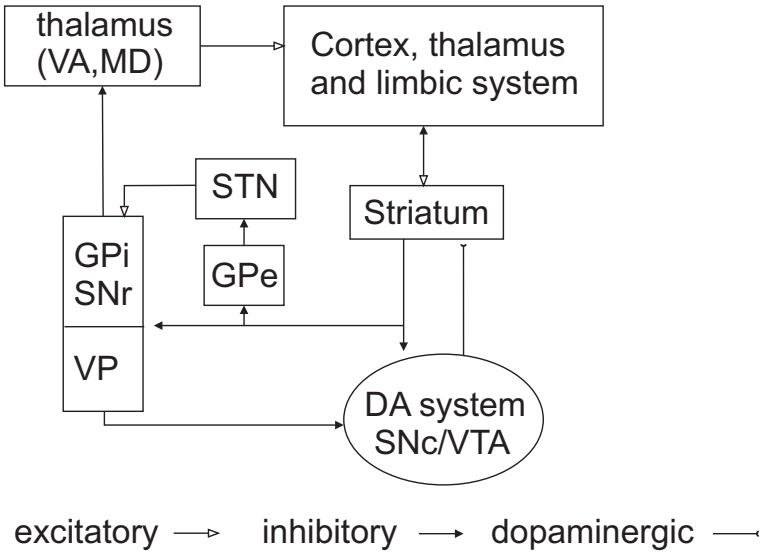

Figure 2.1: A general representation of the basal ganglia-thalamocortical Connectivity. The main input structure to the basal ganglia is the striatum. The striatum is innervated by the cerebral cortex. It sends inhibitory projections to the output nuclei of the basal ganglia including the internal segment of the globus pallidus (GPi) , the substantia nigra pars reticulata (SNr) and the ventral pallidum (VP). These nuclei inhibit the ventral anterior (VA) and mediodorsal (MD) thalamic nuclei. The thalamus is innervated by excitatory inputs from the cortex. The striatum also inhibits the external segment of the globus pallidus (GPe) which inhibits the subthalamic nucleus (STN). The STN sends excitatory projections to the GPi, VP and SNr. (Clark and Boutros, 1999; Joel et al., 2002).

There are two pathways associated with the dorsal division of the basal ganglia namely a direct and an indirect pathway (Alexander and Crutcher, 1990).

The direct pathway's loop originates from the cerebral cortex which project onto the dorsal striatum which innervate the SNr and the GPi that projects onto the thalamus and terminates back to the cortex. The indirect pathway also originates from the cortex which innervate the dorsal striatum. The dorsal striatum projects to the GPe which afferents the subthalamic nucleus. The subthalamic nucleus projects on the GPi which projects to the thalamus and terminates again at the cortex. The subthalamus is the key component

of the indirect pathway. In addition to its efferent projection to the GPi, it also innervates the substantia nigra pars reticulata. The overall effect of activation of the direct and indirect pathway is to increase and decrease cortical activity respectively (Clark and Boutros, 1999). The GPi is a main output of the basal ganglia.

The direct and indirect pathways receive dopaminergic projections from the substantia nigra pars compacta (SNc). These projections make up the nigrostriatal tract (Clark and Boutros, 1999). The direct and indirect pathways are differentiated respectively by the dopamine D1 and D2 receptor types. Dopaminergic fibres act on D1 and D2 dopamine receptors which respectively activate the direct and indirect pathways and indirectly increase and decrease motor activity respectively. Computational models of the basal ganglia have been developed which propose how the basal ganglia performs selection and control through the direct pathway involving D1 receptor activation and indirect pathway involving D2 receptor activation respectively (Gurney et al., 2001a,b, 2004; Prescott et al., 2006).

The rest of this chapter focusses on the ventral striatum. The basal ganglia receive dopaminergic inputs from the SNc and the VTA. While the dopaminergic innervations from the SNc project mainly to the dorsal striatum, the dopaminergic projections from the VTA target the ventral striatum. Fig. 2.2 illustrates approximate projections from the dopaminergic neurons of the VTA and SNc to the ventral and dorsal regions of the striatum, the major input to the basal ganglia. The next section commences with a brief introduction to the vertebrate ventral striatum. It reviews the role of this region in motivation, reward based learning and behavioural flexibility.

2.2 The Ventral Striatum

The ventral striatum, also known as the limbic striatum or nucleus accumbens (NAc), is one of the oldest parts of the brain. This region is particularly interesting because it makes up a critical part of the mesocorticolimbic system

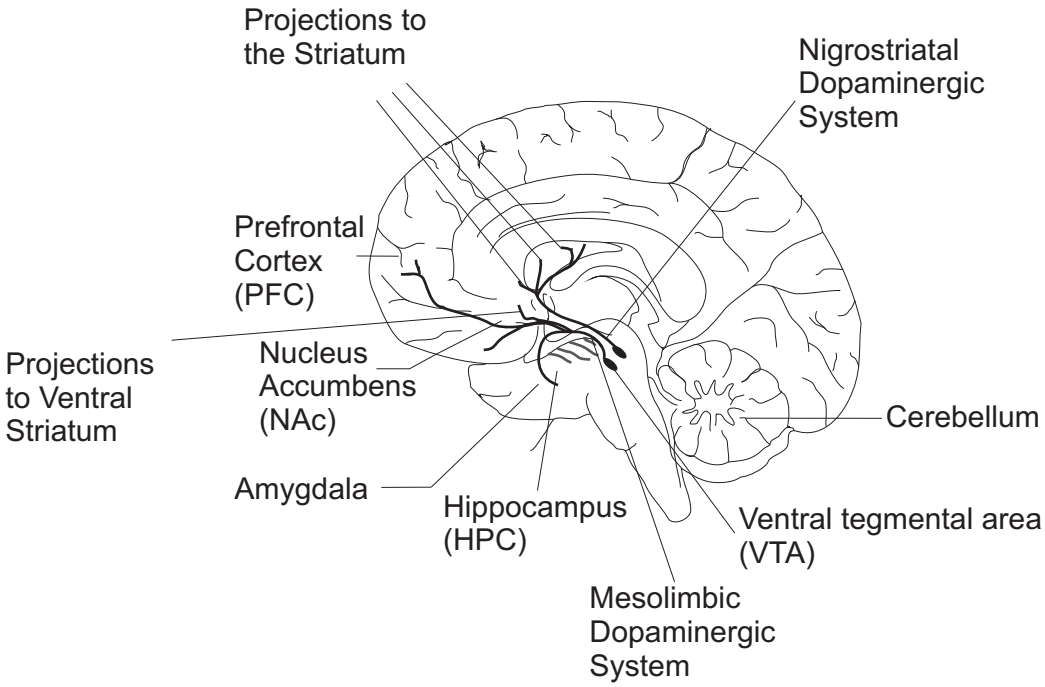

Figure 2.2: An approximate illustration of the projections from the dopaminergic neurons. (Abbreviations: HPC, hippocampus; PFC, prefrontal cortex; VTA, ventral tegmental area; VP, ventral pallidum; NAc, nucleus accumbens).

(Moore and Bloom, 1978) which has been implicated in mediating appetitive learning and reward related behaviours including drinking, feeding, exploration and sex (Kelley, 1999a; Robbins et al., 1989; Robbins and Everitt, 1996). It has also been implicated in the central reward processes associated with electrical brain stimulation (Phillips et al., 1975). In the following sections, the NAc and its surrounding circuitry as well as its role in mediating goal directed behaviours will be elaborated on.

2.2.1 The Nucleus Accumbens (NAc)

The NAc is innervated by limbic structures such as the hippocampus (HPC) (Groenewegen et al., 1987), the basolateral nucleus of the amygdala (BLA)

(Zahm and Brog, 1992), and the medial prefrontal cortex (mPFC) (Zahm and Brog, 1992) and projects to output structures of the basal ganglia such as the ventral pallidum (Robbins and Everitt, 1996). Based on its afferent and efferent structures, the nucleus accumbens integrates information associated with motivation, drive and emotion and translates them into action (Mogenson et al., 1980). Hence the reason for it being identified as the "limbic - motor" interface (Mogenson et al., 1980; Kelley, 1999a). In addition to being innervated by both limbic and cortical regions associated with emotion and cognition respectively, the striatum is also densely innervated by midbrain dopaminergic neurons which originate from the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) (Nicola et al., 2000) and which release the neurotransmitter dopamine (DA).

The NAc can be further dissociated into two anatomically, pharmacologically and behaviourally distinct shell and core subunits (Alheid and Heimer, 1988; Zahm and Brog, 1992; Zahm and Heimer, 1993; Zahm, 2000). Both subunits network in such a way that the core's efferent connectivity resembles that of the dorsal striatum and projects more strongly to basal ganglia regions such as the ventral pallidum and the substantia nigra (Kelley, 1999a) while the shell projects more distinctly to the sub-cortical and limbic regions (Day and Carelli, 2007) including the lateral hypothalamus (LH), ventral tegmental area (VTA) and the ventromedial regions of the ventral pallidum (Kelley, 1999a). Some relevant connectivity surrounding these two subunits are discussed briefly.

2.2.2 The NAc Core

Fig. 2.3 shows some afferent and efferent connectivities the core. The afferent connectivity to this subunit include the amygdala, the dorsal subiculum of the hippocampus (Kelley, 1999a), the dorsolateral part of the ventral pallidum, subthalamic nucleus and the dopaminergic cells of the VTA (Zahm and Brog, 1992). Cortical afferents include the dorsal division of the medial prefrontal cortex (dmPFC) comprising the anterior cingulate which projects

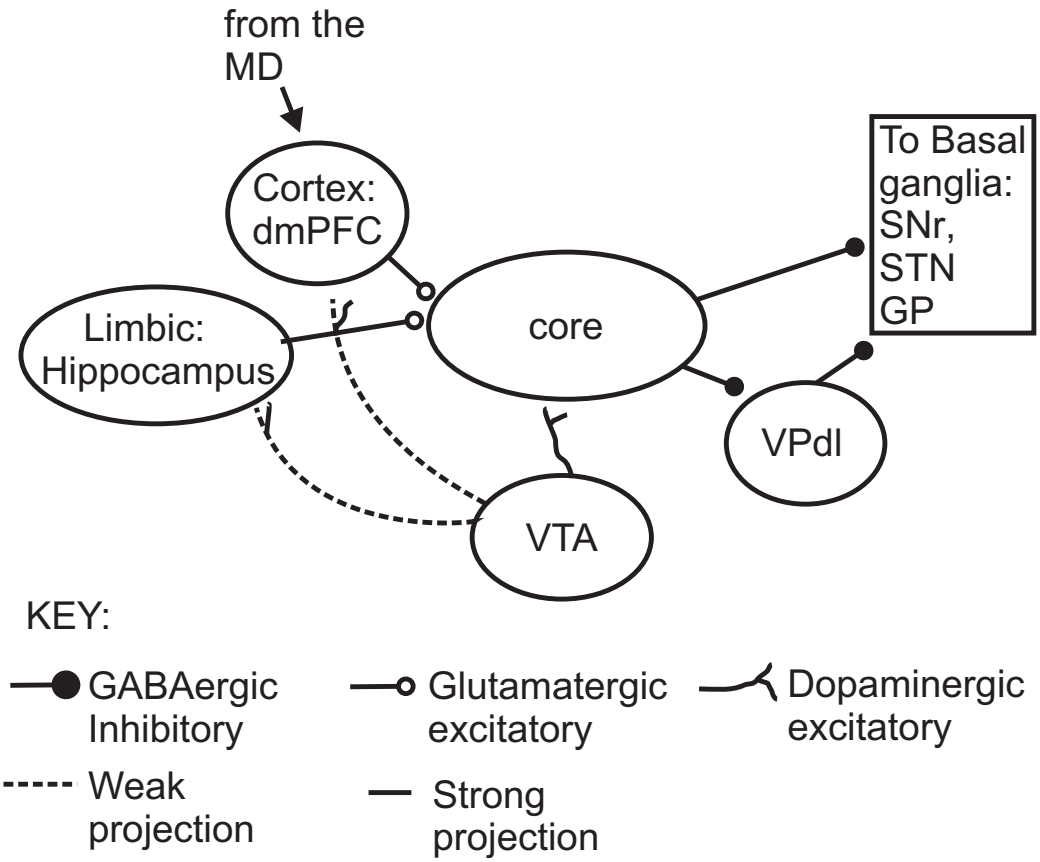

Figure 2.3: A simplified schematic illustrating some afferent and efferent structures that make up the core circuitry. The core receives excitatory glutamatergic innervations from the cortical areas including the dorsomedial prefrontal cortex, and the limbic regions including the hippocampus. The core efferents inhibitory gamma-aminobutyric acid (GABA) innervations to the dorsolateral ventral pallidum, the ventral tegmental area and other basal ganglia structures. (Abbreviations: dmPFC, dorsomedial prefrontal cortex; VTA, ventral tegmental area; VPdl, dorsolateral ventral pallidum; SNr, substantia nigra reticulata; STN, subthalamic nucleus; GP, globus pallidus; MD, mediodorsal nucleus of the thalamus) (Thompson et al., 2009).

more strongly to the core. (Zahm and Brog, 1992; Brog et al., 1993; Passetti et al., 2002). In addition to playing an essential role in working memory, the dmPFC seems to be involved in temporal organisation and shifting of behavioural sequences (Ishikawa et al., 2008).

The efferent connectivity of the core is similar to that of the dorsal striatum and projects more strongly to the output nuclei of the basal ganglia (Zahm and Brog, 1992) via the VPdl. These include the subthalamic nucleus (STN), the substantia nigra reticulata (SNr) and compacta (SNc), the dorsolateral ventral pallidum (VPdl) and globus pallidus (Zahm, 2000).

Based on the similarities of its efferent projections with the dorsal striatum, the core is more associated with the control of voluntary motor functions (Kelley, 2004). The core is necessary for mediating instrumental responding and enables the incentive value of instrumental outcome to control the performance selection. The NAc core enables reward predictive cues to mediate behaviours that led to reward procurement (Kelley, 1999a; Ito et al., 2004). The model has been developed in chapter 3 so that the core representation enables behavior.

2.2.3 The NAc Shell

The connectivity surrounding the shell is shown in Fig. 2.4. The shell is innervated by structures which include the lateral hypothalamus (LH), the ventral subiculum of the hippocampus (Kelley, 1999a), and the medial amygdala (Zahm and Brog, 1992; Ghitza et al., 2003). The hippocampus provides spatial and contextual information to the NAc. The ventromedial prefrontal cortex (vmPFC), which seems to be necessary for maintaining behavioural flexibility of reward based associations (Passetti et al., 2002), comprises the infralimbic and medial orbital cortex which has been suggested to innervate the shell more strongly than the core (Zahm and Brog, 1992; Brog et al., 1993; Zahm, 2000; Passetti et al., 2002; Ishikawa et al., 2008). The shell projects to the VTA, the LH, and the medial part of the ventral pallidum (mVP) (Groenewegen et al., 1999). The shell-mVP connection projects to the VTA (Floresco et al., 2003) and the thalamus (Groenewegen, 1988). The mediodorsal (MD) nucleus of the thalamus projects to the medial frontal cortex (Zahm and Brog, 1992; Birrell and Brown, 2000) which innervates the core. Therefore, the limbic cortico - basal ganglia - thalamocortical circuit in-

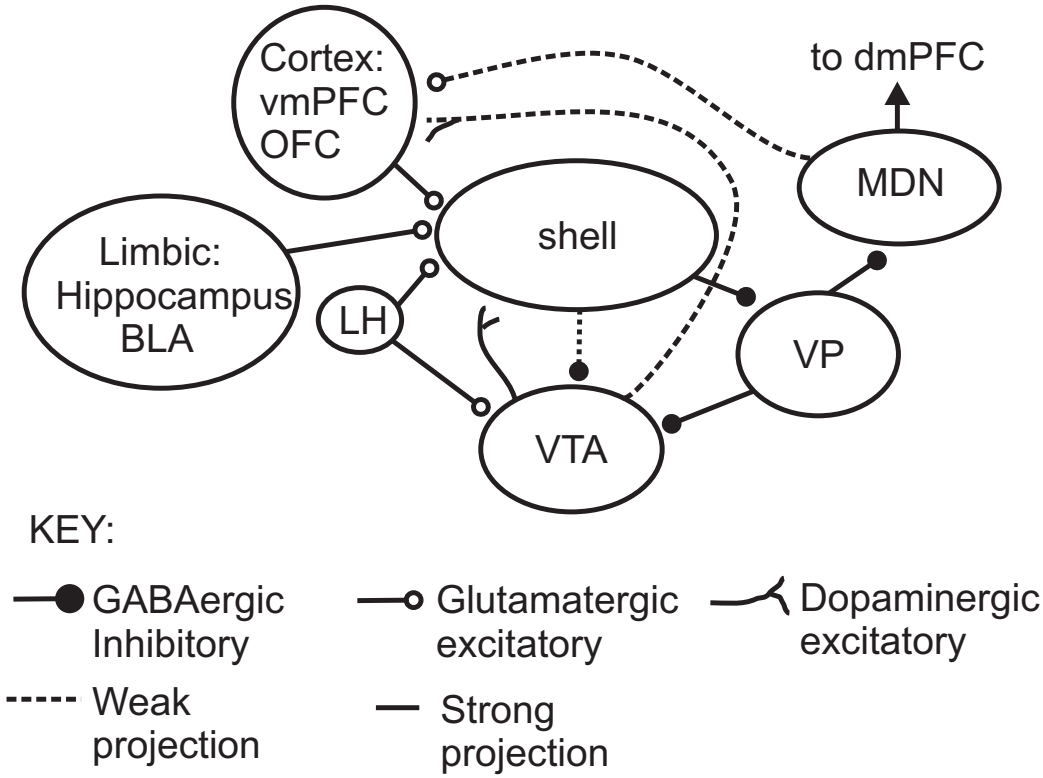

Figure 2.4: A simplified schematic illustrating some afferent and efferent structures that make up the shell circuitry. The shell receives excitatory glutamatergic innervations from the cortical areas including the ventromedial prefrontal cortex and orbitofrontal cortex, the limbic regions including the hippocampus and basolateral amygdala and the lateral hypothalamus. The shell efferents inhibitory GABAergic innervations to the ventral pallidum and the ventral tegmental area. The ventral pallidum sends inhibitory GABAergic projections to the mediodorsal nucleus of the thalamus which feeds excitatory glutamatergic projections back to the cortical regions. (Abbreviations: vmPFC, ventromedial prefrontal cortex; OFC, orbitofrontal cortex; BLA, basolateral amygdala; LH, lateral hypothalamus; VTA, ventral tegmental area; VP, ventral pallidum; MD, mediodorsal nucleus of the thalamus dmPFC, dorsomedial prefrontal cortex) (Thompson et al., 2009).

volving the shell, follows a pathway that leads from the ventral prelimbic and infralimbic cortical areas to the shell to the medial ventral pallidum to the mediodorsal nucleus of the thalamus which then projects back to the cortical areas (Zahm and Heimer, 1990; Groenewegen et al., 1999). It has been sug-

gested by Zahm (2000) that the shell may influence the core activity which could be manifested through this ventral pallido-thalamo-cortical pathway. This connectivity is illustrated in Fig.2.5 which shows how the shell innervated by the cortex, influences the core. The shell inhibits the mVP. The mVP inhibits the MD which projects to the core via the prefrontal cortex. In the computational model, this pathway will be used to suppress unnecessary behaviour initiated by the core activity.

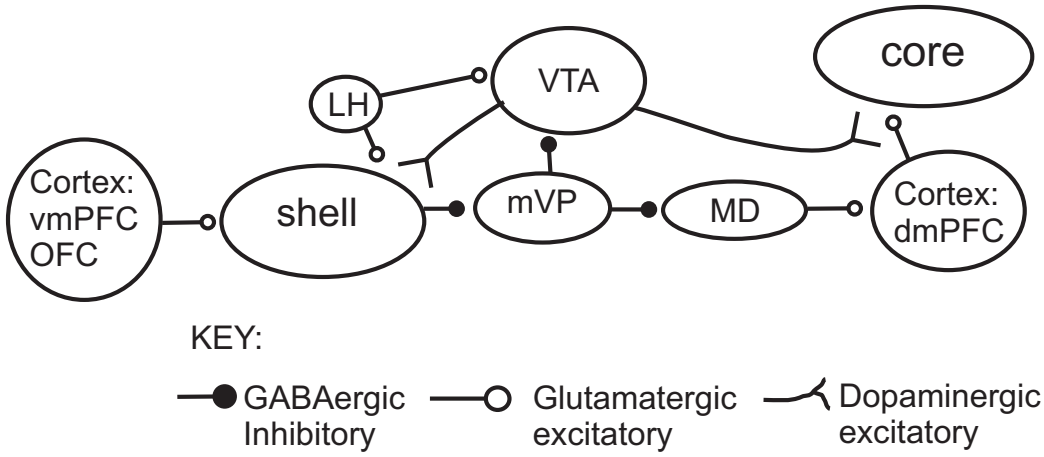

Figure 2.5: A simplified schematic illustrating the essential connectivity between the shell and core implemented in computational model. The shell receives excitatory glutamatergic innervations from the cortical areas including the ventromedial prefrontal cortex and orbitofrontal cortex. The shell inhibits the ventral pallidum which inhibits the mediodorsal thalamus. The mediodorsal thalamus projects to the dorsomedial prefrontal cortex which innervates the core. The shell also influences the ventral tegmental area by inhibiting the vental pallidum which inhibits the ventral tegmental area. (Abbreviations: vmPFC, ventromedial prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; OFC, orbitofrontal cortex; mVP, medial ventral pallidum; LH, lateral hypothalamus; VTA, ventral tegmental area; MD, mediodorsal nucleus of the thalamus).

The innervation from the limbic structures to the NAc are differently modulated by the dopaminergic neurons of the VTA. The NAc has also been observed to influence DA release (Floresco et al., 2003). This means that the limbic structures innervating the NAc can indirectly influence dopamine re-

lease. The NAc and the DA neurons of the VTA are innervated by excitatory glutamatergic neurons of the lateral hypothalamus which can be activated by primary rewards. Manipulating the DA receptors associated with the NAc target structure have demonstrated different adjustments on rewarding effects (Phillips et al., 1994).

There are two main DA systems (Fig. 2.6) namely, the mesocorticolimbic-DA system originating from VTA neurons and innervating the nucleus accumbens (NAc) and cortical and limbic structures, and the nigrostriatal (NS) dopaminergic system originating from the substantia nigra compacta (SNc). In the next section dopamine and the mesocorticolimbic DA system are briefly discussed.

2.3 The Mesocorticolimbic Dopaminergic System

DA was first discovered in the late nineteen fifties by Arvid Carlsson (Carlsson and Waldeck, 1958; Abbott, 2007) and has been recognised as an important neurotransmitter in the reward circuitry of the brain (Wise and Rompre, 1989). It plays a role in both synaptic plasticity and memory processes. Dopaminergic (DAergic) activity on the NAc has been implicated in a variety of cognitive, motivational and behavioural functions such as motor activity (Dalia et al., 1998), responding to salient or novel stimuli (Rebec et al., 1997; Horvitz, 2000) and has also been observed to affect locomotion which can be blocked by disabling the connectivity of the NAc efferents to the globus pallidus (Jones and Mogenson, 1980). It has been suggested by (Mogenson et al., 1988) that DA instead, modulates the effects of afferent inputs to the NAc. Studies have shown that the activation and inactivation of the NAc DA systems respectively generated increased and decreased behavioural responses to reward predictive cues towards obtaining goal objects (Wise, 1998). DA is essential in mediating or gating and as such, in selecting information from the limbic and cortical regions that innervate and activate

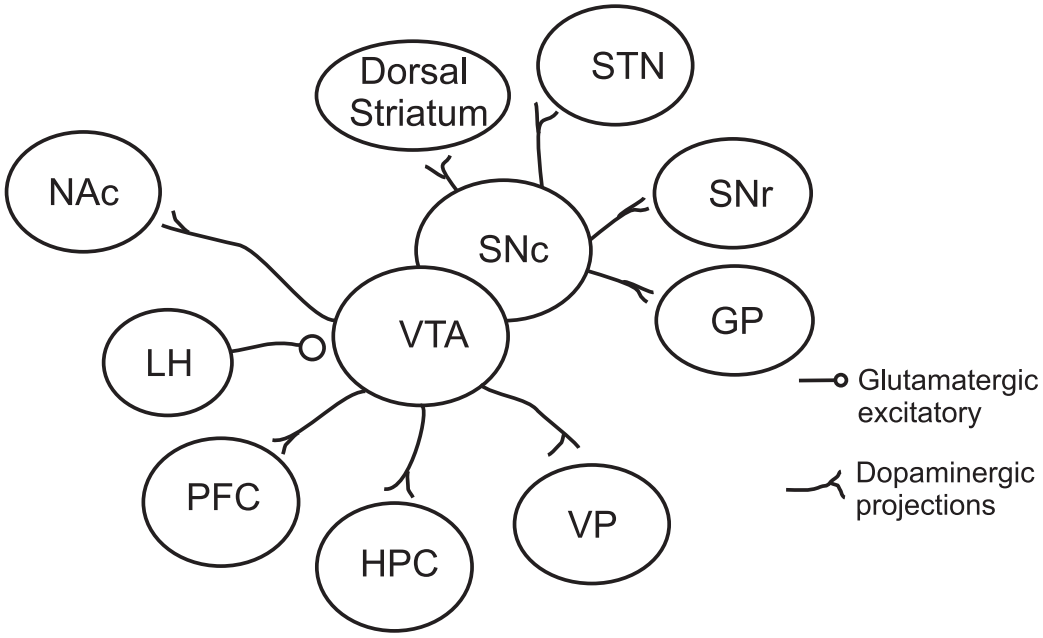

Figure 2.6: The two dopaminergic systems. The mesocorticolimbic DA system originates from the VTA and innervates structures which include the nucleus accumbens, lateral hypothalamus, prefrontal cortex, hippocampus and ventral pallidum. The nigrostriatal DA system originates from the substantia nigra compacta and innervates regions including the dorsal striatum, the subthalamic nucleus, the substantia nigra reticulata and the globus pallidus. (Abbreviations: NAc, nucleus accumbens; LH, lateral hypothalamus; PFC, prefrontal cortex; HPC, hippocampus; VTA, ventral tegmental area; VP, ventral pallidum; MD, mediodorsal nucleus of the thalamus; STN, subthalamic nucleus; SNr, substantia nigra reticulata; SNc substantia nigra compacta GP, globus pallidus) (Thompson et al., 2009).

the NAc (Cepeda et al., 1998).

The discovery of intracranial self stimulation in 1954 (Olds and Milner, 1954) led to studies which have shown that DA plays a primary role in mediating reward related and goal directed behaviours (Wise, 1998, 2004). The two main DA systems (Fig. 2.2) include the mesocorticolimbic-DA system originating from VTA neurons and innervating the nucleus accumbens (NAc) and the nigrostriatal (NS) dopaminergic system originating from the substantia nigra pars compacta. The focus will be on the mesocorticolimbic-DA system

because it has been identified to play more of a major role in motivation and reward functions than the NS DA system (Alcaro et al., 2007; Papp and Bal, 1987).

DA neurons exhibit burst spiking activity in receipt of primary rewards such as food, novel appetitive stimulus and in event of stimulus which predict rewards (Steinfels et al., 1983; Romo and Schultz, 1990; Ljungberg et al., 1992; Schultz et al., 1993, 1997). When cues are paired repeatedly with the rewards, these DA activations develop more significantly at the onset of the cue presentation and less during reward delivery (Schultz et al., 1997).

DA neurons are innervated by the excitatory glutamatergic projections from the lateral hypothalamus (LH) and inhibitory GABAergic afferents from the NAc and ventral pallidum (VP). The VTA-DA neurons exhibit two transmission modes namely phasic and tonic activity described as follows.

2.3.1 The Spiking Activity of DA Cells

According to physiological findings, the phasic and tonic levels of DA release are dependent on the two distinct methods that drive the spiking activity of the VTA DA neurons (Fig. 2.7). Burst firing of DA neurons at an approximate frequency of 3Hz generate phasic DA levels in the synaptic cleft which are very quickly removed by dopamine transporters (Floresco et al., 2003; Grace et al., 2007) while tonic DA levels occur in the extrasynaptic space at extremely low levels due to an increase in the number of tonically active DA neurons (Fig. 2.7). Floresco et al. (2003) observed that VTA-DA increase can occur via glutamatergic excitations or GABAergic dis-inhibition. When primary rewards are obtained, the LH, which sends excitatory glutamatergic inputs to the DA cells, becomes activated. VTA-DA cells demonstrate burst firing in response to behaviourally relevant stimuli such as rewards (Schultz et al., 1997) which can occur due to the VTA's innervation by the LH glutamatergic projections. It is believed that these burst firing activity signal reward useful for goal directed behaviour (Schultz, 1998; Grace et al., 2007).

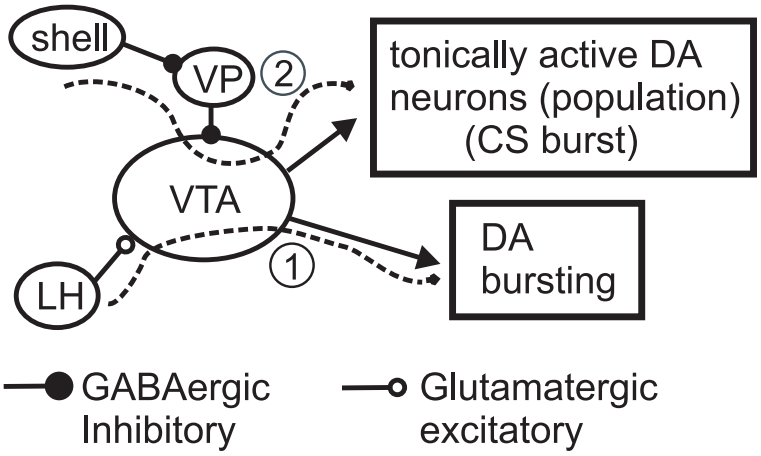

Figure 2.7: The spiking activity of DA neurons influenced by (1) a direct excitatory pathway and (2) a dis-inhibition of the ventral pallidum neurons. (Abbreviations: LH, lateral hypothalamus; VTA, ventral tegmental area; VP, ventral pallidum) (Thompson et al., 2009).

The VTA is also innervated by inhibitory GABAergic projections from the NAc and VP. Activation of the NAc produces an inactivation of the inhibitory GABAergic VP-VTA projections and a resultant increase in the population activity of DA neurons (Floresco et al., 2003). Therefore, LH-glutamatergic excitation generates burst spiking at the moment of the primary reward while the NAc-VP-GABAergic dis-inhibition is responsible for tonic levels of DA.

In this section two DA activity modes have been described. In the following sections, the role these DA activities play on signalling and synaptic transmission in the NAc and striatum will be investigated and discussed.

2.4 Signalling and Synaptic Transmission

Long term synaptic efficacy of excitatory signalling are proposed to be essential mechanisms necessary for the storage of information and the cellular basis for memory, learning, motor control and different behaviours (Calabresi et al., 1997; Gubellini et al., 2004; Schotanus and Chergui, 2008). While sig-

nalling and spiking activity in the NAc and striatum are influenced by the corresponding limbic and cortical afferent structures, studies observing DAergic mechanisms in modulating synaptic transmission have generated rather contradictory results (Nicola et al., 2000). This section investigates briefly, some of these studies and summarises certain procedures that take place during signalling in the striatum and the NAc.

A vast majority ($> 90\%$) of NAc neurons are composed of inhibitory GABAergic medium spiny neurons MSNs while the rest make up three groups of interneurons (Groves, 1983; Meredith, 1999; Calabresi et al., 2007) listed as follows:

1. Slow-firing cholinergic interneurons which co-express D1-type and D2-type receptors (Nicola et al., 2000).
2. Fast-spiking paravalbumin-containing GABAergic interneurons.
3. Burst-firing somatostatin/nitric-oxide (NO) expressing interneurons (Nicola et al., 2000; Calabresi et al., 2007).

Neurotransmitter and Receptor Activity

Cortical regions such as the PFC release excitatory glutamate (Glu) neurotransmitter (Divac et al., 1977) on to the NAc of the ventral striatum and form cortico-striatal synapses. A proportion of striatal neurons synapse with both dopaminergic and cortical innervations (Freund et al., 1984) where glutamatergic and dopaminergic interactions occur (Smith and Bolam, 1990; Day and Carelli, 2007) and act on receptors located pre- and postsynaptically (Reynolds and Wickens, 2002). Interactions between glutamate and DA from afferent innervations form part of the mechanisms necessary for signalling and striatal synaptic activities. Postsynaptic depolarizations occur on the NAc or striatum due to the excitatory glutamatergic activations from limbic and cortical structures which bind to different glutamate subtypes including ionotrophic N- methyl- D- aspartic acid (NMDA) and α-

amino- 3- hydroxyl- 5- methyl- 4- isoxazole- propionate (AMPA) and other metabotropic type glutamatergic (mGlu) receptors (Nicola et al., 2000). In vivo the MSNs, often called bimodal neurons, exhibit fluctuations in their membrane potential between two functional states. A negative and quiescent hyperpolarized downstate ($\simeq 80mV$) and a depolarized upstate ($\simeq 60mV$) (Tseng et al., 2007). Depolarization of these MSNs into upstate have been observed to require the activation of AMPA and/or NMDA receptors (Tseng et al., 2007) which can be achieved by strong glutamatergic inputs.

DA is a neuromodulator of striatal synaptic plasticity that acts on corresponding DA receptors. There are five DA receptors D1-D5 that have been cloned. Due to their molecular and pharmacological similarities, they can be grouped into two family subtypes (Nicola et al., 2000; Calabresi et al., 2000). The D1-type family of receptors represent D1 and D5 DA receptors, while D2, D3 and D4 receptors make up the D2-type family of DA receptors. These DA receptors are also located on striatal neurons and its afferent innervations. In particular, most D1 receptors are localized postsynaptically on the NAc's MSNs while D2 receptors are located presynaptically on corticostriatal fibres, on DAergic neuron terminals and also postsynaptically on MSNs (Calabresi et al., 1997; Kötter, 1994).

The state of this membrane potential of MSNs can determine whether the resultant dopaminergic activation of DA receptors are excitatory or inhibitory (Reynolds and Wickens, 2002). These postsynaptic depolarizations are also referred to as excitatory postsynaptic potentials (EPSPs). On the other hand, postsynaptic hyperpolarizations or inhibitory post-synaptic potentials (IPSPs) are generated when the inhibitory neurotransmitters such as GABA are released on the NAc or striatum.

Dopamine Signalling and Transmission

Manipulation of DA receptors associated with the striatal circuitry has demonstrated different adjustments on rewarding effects (Phillips et al., 1994). The distribution of D1 and D2 receptors in the striatum appears to be largely

segregated such that the neurons involved in the direct and indirect pathways express high levels of D1 and D2 receptors respectively. In addition, a certain population ($\simeq 25\%$) of medium spiny neurons of the striatum have also been observed to co-express these two subtypes of receptors. In the NAc, the exact amount of colocalization of D1- and D2-type receptors is debatable since although D1 and D2 receptors do not seem to be expressed together, D3, receptors are expressed in both regions of D1 and D2 receptor expression (Nicola et al., 2000). The activations of these D1- and D2-type receptors depend on the transmission states of the DA neurons and enables the long term increase (long term potentiation, LTP) or decrease (long term depression, LTD) in the strength of the striatal synapses.

Phasic and tonic DA activity which generate different DA concentration levels have been suggested to play a role in mediating LTP and LTD. DAergic activations of D1- and D2-type receptors influence corticostriatal synaptic plasticity (Calabresi et al., 2007). The function DA receptors play on corticostriatal synaptic transmission (Pawlak and Kerr, 2008; Surmeier et al., 2007) is dependent on DA concentration (Hsu et al., 1995). According to Garris et al. (1994), phasic DA is released in concentrations in the order of hundreds of micromolars (Grace, 2000). In contrast, tonic extracellular DA levels in the NAc occur at concentrations in the range of nanomolars ($5 - 50nm$) (Parsons and Justice, 1992; Grace, 2000). Such low DA concentrations ($\leq 0.1\mu M$) act on D2 receptors (Pawlak and Kerr, 2008). While this D2 receptor activation does not seem to be essential for timed pre- and postsynaptic activity to induce plasticity (Pawlak and Kerr, 2008), D2 receptor stimulation act on the presynaptic mechanisms involved in the release of glutamate (Lee et al., 2005). At higher concentrations ($\geq 0.1\mu M$), both postsynaptic D1 and D2 receptors are activated which may also result in an attenuated synaptic transmission (Lee et al., 2005).

Dopamine activation on these DA receptors have been observed to have both excitatory and inhibitory effects on striatal synaptic plasticity (Mercuri et al., 1985) depending on the DA receptor subtype activated (Nicola et al., 2000). However, the effect of dopamine on striatal neurons can also be dependent on

the membrane potential of the postsynaptic neuron. For example, D1 receptor activation during postsynaptic depolarisation has been observed to enhance the effect of excitatory inputs (Reynolds and Wickens, 2002; Surmeier et al., 2007). With cortical and dopaminergic terminals occurring proximally on striatal neurons, certain forms of synaptic plasticity in the striatum have been suggested to be induced by three factors (Kötter, 1994; Reynolds and Wickens, 2002; Porr and Wörgötter, 2007). These include both glutamatergic activation and depotentiation of the pre- and postsynaptic activities respectively and DA modulation as the third factor. DA modulation occurs in two activity states and differentially act on the DA receptors. Accordingly DA activity functions as a gate and enables an increase or decrease in synaptic plasticity. Therefore, the phasic and tonic DA modes acting differentially on DA receptors can influence the plasticity of corticostriatal synapses. The methods by which phasic and tonic DA facilitate LTP and LTD respectively, will be used in the computational model developed in chapter 3.

The synapses that link afferent units to the striatum can undergo long term plastic changes such as LTP and LTD (Calabresi et al., 1992c). These changes, which are likely to influence learning and memory formation (Calabresi et al., 1996), occur within milliseconds but can last for hours and even days (Di Filippo et al., 2009). DA denervation has resulted in impaired synaptic plasticity (Picconi et al., 2005). Two different activity states of DA have been briefly discussed so far which result in distinct DA concentration levels that play a role in mediating synaptic plasticity by stimulating DA receptors. The following section discusses the possible role dopamine plays in mediating long term depression at corticostriatal synapses.

2.4.1 Long Term Depression (LTD)

Long term depression (LTD) and long term potentiation (LTP) are two main forms of synaptic plasticity which have been observed at striatal synapses (Calabresi et al., 2007). Although there are studies that report no observable effects of DA on striatal synapses (Nicola et al., 2000), a few studies which

have observed LTD as well as DA's involvement in striatal signalling or LTD are summarised in table 2.1 and discussed as follows:

(HFS) of corticostriatal fibres on the synaptic function of striatal neurons, Calabresi et al. (1992a) observed that HFS of corticostriatal glutamatergic fibres in the striatum induced LTD which were blocked by either D1 or D2 receptor antagonists. In addition LTD was abolished by DA depletions and restored either by the applications of DA or administration of D1 and D2 receptor agonists (Picconi et al., 2005). Studies conducted by Cepeda et al. (1993); Hsu et al. (1995); Levine et al. (1996) show that D2-type receptor activations attenuate EPSPs mediated by Glu receptors, Therefore it can be assumed that D2 receptor activation have inhibitory effects on the NAc.

D2 receptor activation seems to be an essential requirement for the induction of LTD. This is because a failure to demonstrate LTD has been noted in D2 deficient mice. In addition, mice lacking the Dj-1 gene which exhibits reduced DA overflow in the extrasynaptic striatal spaces also showed failed LTD induction (Calabresi et al., 2007). Similar results were observed during bath application of DA on striatal synapses (Calabresi et al., 1992b, 1987). Both DA overflow in the extrasynaptic space and bathing DA simulate tonic DA levels. Cortico-accumbens transmissions appear to be attenuated in presence of tonic DA levels (O'Donnell and Grace, 1994; Floresco et al., 2003; Goto and Grace, 2005a). According to Maeno (1982) and Creese et al. (1983) D2 receptors show a high affinity for DA and could be activated in the event of tonic DA release (Goto and Grace, 2005a; Grace, 1991). Therefore tonic DA production via the inactivation of the VP have resulted in the selective attenuation of mPFC afferents to the NAc (Goto and Grace, 2005b; Grace et al., 2007). These findings demonstrate that LTD can be induced and blocked by the activation of D1 and D2 receptors and the application of D1 and D2 receptor antagonist respectively (Calabresi et al., 1992a,b).

Certain studies exist which aim to explain the mechanisms involved in LTD induction. These processes include a variety of factors that involve a chaining of internal events. According to Gubellini et al. (2004), during corti-

Table 2.1: Table showing DAergic influences on striatal synaptic plasticity

Reference	Manipulation	Effect	Comment
Calabresi et al. (1992a)	Corticostriatal LTD by HFS	· Blocked by D1 or D2 R antagonists · abolished by DA depletion · Restored by DA application or · by D1 and D2 R agonists	DA acting on D1 and or D2 R are involved in initiating corticostriatal LTD
Gonon and Sundstrom (1996) and Gonon (1997)	Excitation of NAc neurons	· Abolished by DA lesioning · Reduced by D1 R antagonist · Facilitated by D1 R agonists	DA and D1 R activation are involved in facilitating excitatory stimulation of the NAc
Schotanus and Chergui (2008)	Corticostriatal LTP by HFS	· Blocked by D1 R antagonist · Blocked by excess DA in the extracellular space.	DA acting on D1 R and not D2 R are involved in inducing corticostriatal LTP

costriatal HFS, glutamatergic AMPA receptor stimulation which results in action potential discharges are essential for LTD induction. The postsynaptic depolarizations results in increased intracellular calcium (Ca^{2+}) which trigger several Ca^{2+}- dependent processes that are required for LTD induction. Postsynaptic depolarizations which result in increased Ca^{2+} levels include one mechanism required for LTD induction.

Another mechanism implicated in LTD induction, requires DA receptor stimulation and has been discussed by both Wang et al. (2006) and Calabresi et al. (2007). It involves one of the interneurons located in the striatum, the cholinergic interneuron which expresses D2 receptors (Wang et al., 2006). DA acting on the D2 receptors generate reduced activity of muscarinic M1 acetylcholine receptors. Activation of M1 receptors generally results in an attenuation of the opening of Cav 1.3 Ca^{2+} channel (Wang et al., 2006) and therefore reduced endocannabinoid (ECB) synthesis and release. LTD induction seems to essentially require ECB release. Postsynaptic endocannabinoids signal the activation of presynaptic cannabinoid CB1 receptors which are essential for the reduction of glutamate release and thus the initiation of LTD (Yin et al., 2006). Therefore D2 receptor activation reduces M1 receptor activation which indirectly enhances ECB release required for LTD induction. A third mechanism briefly addressed here includes a process also suggested by Calabresi et al. (2007) and involves the nitric oxide producing interneuron which contain D1-type receptors. In this case, burst firing of DA neurons activate the D1-class receptors located on this interneuron which in turn release NO. This mechanism has been suggested to influence the induction of LTD (Calabresi et al., 2007). LTD induction seems to require a chain of subcellular processes which are triggered by certain events including presynaptic glutamate and dopamine release which act on their respective receptors located pre- and postsynaptically as well as on interneurons.

It has been proposed by Calabresi et al. (1996) and Law-Tho et al. (1995) that in the PFC, LTD is favoured over LTP in the presence of DA. It is assumed that this mechanism might also occur at corticostriatal sysnapes. This leads to the suggestion that tonic DA-D2 receptor stimulation might

enable corticostriatal LTD to occur. However, D1 receptor activation might initially be required to induce synaptic plasticity. A combination of pre- and post synaptic activities have been observed to induce LTD with a possible dependence on the presence of D2 receptors (Reynolds and Wickens, 2002).

To summarize, LTD induction requires presynaptic glutamate release that act on glutamate receptors and the activation of DA D1 and/or D2 receptors. This will be used in the development of the model in the next chapter. In the model, it will be assumed that tonic DA activation of D2 receptors mediated LTD. Some studies which implicate dopamine activity on striatal signalling and synaptic plasticity are summarised in table 2.1. In the following section, it will be discussed how LTP is induced.

2.4.2 Long Term Potentiation (LTP)

Striatal LTP can be induced after the stimulation of cortical and hippocampal afferents (Pennartz et al., 1993; Boeijinga et al., 1993) in particular, corticostriatal LTP has been observed during coincident tetanic stimulation of corticostriatal and glutamatergic NMDA receptor activation. Glutamate release acting on NMDA receptors are essential for the induction of LTP. In the presence of magnesium (Mg^{2+}), NMDA receptor channels become inactivated and LTP induction fails due to the voltage dependent Mg^{2+} blockade (Calabresi et al., 1992c). This voltage dependent Mg^{2+} blockade occurs due to a very negative resting membrane potential of striatal neurons and can be overcome by membrane depolarization. Repetitive HFS of the corticostriatal pathway rather than individual single unit activation might be essential in the removal of this Mg^{2+} blockade (Calabresi et al., 2000). In addition, Charpier and Deniau (1997) have demonstrated LTP induction during combined repetitive activation of corticostriatal neurons and striatal depolarization. Thus striatal LTP occurs in event of combined pre- and postsynaptic activities.

In addition to the combined pre- and postsynaptic mechanisms, the HFS induction of corticostriatal LTP has been observed to require the activation of

DA D1 receptors (Calabresi et al., 1992b, 2000; Kerr and Wickens, 2001). In the NAc, LTP has been induced by HFS of glutamatergic inputs acting on NMDA receptors, which in turn has been blocked by D1 receptor inactivation or excess DA in the extracellular space (Schotanus and Chergui, 2008) (summarised in table 2.1). While LTP has been induced in event of corticostriatal activation, dopamine depletion at corticostriatal synapses (Centonze et al., 1999) or a blockade of D1 type receptor activation (Calabresi et al., 1992c; Kerr and Wickens, 2001) have both resulted in a failure to induce LTP. Additional studies have shown that glutamate mediated EPSPs are enhanced by D1 receptor stimulations (Galarraga et al., 1997; Umemiya and Raymond, 1997). DA concentration levels act on corresponding DA receptors which in turn can induce synaptic plasticity. The importance of D1 receptor activation in producing LTP is consistent with the finding that LTP induction is absent in mice lacking D1 receptors (Calabresi et al., 2007). Corticostriatal LTP can occur in event of pre- and postsynaptic activities when the essential DA D1-type receptors are stimulated (Calabresi et al., 2000; Kerr and Wickens, 2001). D1 receptors can be activated when unexpected rewards generate burst spike firing and phasic DA release (Grace et al., 2007; Goto and Grace, 2008).

2.4.3 The Interplay between LTP and LTD

Pawlak and Kerr (2008) demonstrated that although both an increase and decrease in synaptic plasticity can be generated under similar conditions, the timing between the presynaptic cortical inputs and the postsynaptic striatal activations also play a role in inducing LTP and LTD. In addition, the D1 type receptor activation was necessary for both forms of striatal spike timing dependent plasticity (STDP) while D2 receptor blockade resulted in enhanced and delayed potentiation and depression respectively.

LTP and LTD seem to require distinct processes involving DA receptor activation. D2 receptor activation and blockade observed to respectively disrupt and enhance LTP, led to the suggestion by Calabresi et al. (2007) that unlike

in the induction of LTD, D1 and D2 dopamine receptors operate in opposition to induce LTP. To summarize, LTP induction requires three elements which include a precise timing between the first two elements, pre- and postsynaptic activities and DA receptor stimulation as the third factor. D1 receptor activation mediates the induction of both LTP and LTD, D2 receptor activation seems to favour the induction of LTD.

DA receptors activation can be influenced by different spiking activity states of DA neurons. Phasic DA release generated by VTA burst firing (which can occur when rewards are obtained) activates D1 receptors (Goto et al., 2007). Alternatively, tonic DA release generated from the dis-inhibition of the VTA activate D2 receptors (Goto et al., 2007). In addition to influencing synaptic plasticity in the NAc, Phasic and tonic DA levels have been proposed to facilitate limbic and cortical inputs through the activation of the D1 and D2 receptors respectively (Goto and Grace, 2005b).

The NAc receives inputs from cortical and limbic regions and DAergic neurons. These afferent innervations undergo synaptic plasticity which make up the cellular basis for behaviour and motor learning. So far studies have been identified which implicate mainly DA in corticostriatal signalling and synaptic plasticity. The following sections observe how manipulating afferent influences including DAergic and glutamatergic innervations influence behaviour in the striatum and NAc. In addition several lesion and inactivation studies will also be summarised and used to suggest how signalling and information is transferred at a systems level through the mesocorticolimbic circuitry to influence and regulate motivation, reward and goal directed behaviours.

2.5 Experimental Studies of the NAc Circuitry and Functionality

Although there are numerous experimental studies which demonstrate that the NAc plays a role in motivation, rewards and goal directed learning (Kelley, 1999a; Corbit et al., 2001; Reynolds and Berridge, 2001; Yin et al., 2008; Nicola et al., 2004; Ikemoto and Panksepp, 1999) by processing information from the limbic, sub-cortical and dopaminergic inputs, the exact mechanisms involved are still uncertain. However, by studying the experimental manipulations of the dopaminergic and afferent innervations involved with the NAc, an understanding of the role of the NAc in action selection, motivation and goal directed behaviour can to some extent, be realized.

Experimental studies have shown impaired response to cues that predict reward due to a reduction of DA activity on the NAc (Di Ciano et al., 2001; Parkinson et al., 2002; Robbins et al., 1989; Nicola, 2007; Wakabayashi et al., 2004; Yun et al., 2004). Lesion studies have demonstrated disruptions of processes which range from cognitive to behavioural flexibility. There exists a vast range of literature which implicate the DA and the mesocorticolimbic system in animal learning and rewarding behaviours such as Pavlovian or classical conditioning and instrumental conditioning. Classical and instrumental conditioning in the NAc are summarised in the following section.

2.5.1 The NAc in Pavlovian and Instrumental Mechanisms

Pavlovian and instrumental conditioning as introduced in chapter 1, are elementary forms of associative learning which allow animals to predict and adapt to changes in their environment based on their previous experiences. Pavlovian conditioning affects behaviour and includes mechanisms such as autoshaping, conditioned reinforcement and Pavlovian instrumental transfer (PIT) (Cardinal et al., 2002a). Classical and instrumental conditioning

have been observed in the brain regions including the cerebellum which has been associated with the nictitating membrane (Welsh and Harvey, 1989) and the amygdala which is involved in fear conditioning (Park and Choi, 2010). However, a number of studies have implicated the NAc as a relevant region involved in modulating the influence of appetitive Pavlovian CSs on instrumental behaviour (Parkinson et al., 2000, 2002). These processes along with experimental manipulations on the NAc circuitry can be used to demonstrate how reward predictive cues and stimuli are implemented by the NAc to influence reward based and goal directed behaviours.

Pavlovian Conditioned Approach Behaviors

In sign-tracking or autoshaping, when a conditioned stimulus (CS) has been associated with an unconditioned stimulus (US), a conditioned response (CR) in the form of an approach behaviour towards the CS is elicited irrespective of the presentation of the US (Brown and Jenkins, 1968). An example of autoshaping is demonstrated when a hungry pigeon is placed in a conditioning chamber in which there is contained a perspex which becomes illuminated shortly before food is delivered in a food hopper. Initially the pigeon does not respond to the illuminated panel however, after a few trials, the pigeon will come to peck the panel when it is lit. The food delivery is not dependent on the pigeons response therefore it can be classified as an open-loop system. Goal-tracking is similar to sign-tracking however, approach behaviour elicited is in the direction of the site of the US (Silva et al., 1992). Sign- and goal-tracking which are approach behaviours in response to a CS towards the CS or US respectively demonstrate the acquisition of Pavlovian conditioned approach (PCA) behaviour.

Lesion and inactivation experiments have shown that the NAc plays a role in autoshaping. For instance, glutamatergic and dopaminergic receptor inactivations (Di Ciano et al., 2001) as well as lesions to the core and not the shell (Parkinson et al., 2000) have been observed to disrupt acquisition and performance in discriminated Pavlovian approach behaviours. In particular,

application of DA receptor antagonists on the NAc core exhibited decreased approaches to the CS paired with a reward (CS+) during both acquisition and performance of the conditioned responses (Di Ciano et al., 2001). A blockade of NMDA receptors in the core demonstrated impaired acquisition of autoshaped response while AMPA receptor blockade demonstrated increased approaches to the CS that was not paired with the reward (CS-). DA depletion on the NAc also showed impaired acquisition and response to Pavlovian autoshaped behaviour (Cardinal et al., 2002b; Parkinson et al., 2002). These studies confirm that DA and glutamate transmission in the core are both involved in the acquisition of CS-US associations. In addition impaired autoshaping demonstrated specifically by core and not shell lesions implies that there seems to be a dissociation in the processes by which the NAc's two subregions function. The core seems to be a site for which Pavlovian CSs signalled by glutamatergic afferents mediate instrumental response which is facilitated by DA transmission. Additional studies which affirm the NAc's role in influencing behavioural responses can also be observed in instrumental conditioning (Dickinson et al., 2000; Corbit et al., 2001; Hall et al., 2001; Lex and Hauber, 2008).

In experiments observing the effects of the shell and core inactivations, in goal tracking PCA behaviours, Blaiss and Janak (2008) found that the NAc was necessary for the expression and not the consolidation of goal tracking PCA. In particular, the core is essential in the expression of CS-US associations, while the shell plays a role in inhibiting conditioned approach behaviour when the CS that precedes a reward is omitted.

Instrumental Conditioning and Outcome Devaluation

During instrumental conditioning, rewards are delivered after a certain response has been made. Therefore instrumental conditioning can be identified as a closed-loop system. Corbit et al. (2001) performed experiments to observe the effects of NAc shell and core lesions in instrumental conditioning. Core lesioned animals performed at a generally lower response rate than

shell or sham lesioned agents during both training in outcome devaluation experiments and contingency degradation experiments. While core lesioned groups did not demonstrate sensitivity to devalued outcomes, they showed some form (although reduced) of discrimination towards degraded instrumental action-outcome contingencies. These findings suggest that the core is involved in initiating and controlling actions depending on the reward value of the outcome (Corbit et al., 2001). In other words, the core seems to be responsible for enabling actions which have rewarding outcomes. In addition to instrumental learning, the NAc has also been implicated in transfer effects of Pavlovian CSs on instrumental responses.

Pavlovian Instrumental Transfer

Pavlovian instrumental transfer PIT consists of two phases, a Pavlovian phase and an instrumental phase. During the Pavlovian phase, agents are placed in the operant chambers and are presented with two discriminative stimuli a CS+ and a CS-. The CS+ is paired with a reward while the CS- does not produce an output. During presentation of the CS+, the agent learns to approach the location where the reward is delivered. In the instrumental phase the agents are presented with two levers. A response to the active lever results in the delivery of the reward while a response on the second lever has no consequence. The agent learns to respond on the active lever. In the test phase the CS+ and CS- are presented without any rewards delivered and the degree of responding by the agents is observed. The agents demonstrate enhanced responding to the active lever during the CS+ compared to the CS-. Therefore PIT consists of a training whereby a CS is paired with a US and then the CS is presented while the subject performs an instrumental process (Pearce, 2008). It involves a process by which a stimulus previously paired with a reward (CS+), enhances instrumental responding (Cardinal et al., 2002a).

While examining the effects of NAc shell and core lesions in instrumental learning, Corbit et al. (2001) observed that PIT was completely eliminated

by shell lesions. The shell seems to be essential in transforming or mediating information from reward predictive cues to enabling instrumental responding necessary for obtaining rewards. On the other hand core rather than shell lesions were observed to disrupt PIT (Hall et al., 2001). Therefore both core and shell lesions have been implicated in the general form of PIT in particular, the shell seems to be a structure through which reward related cues have an excitatory effect on goal directed instrumental processes (Corbit et al., 2001).

DA receptor activation on the NAc also seems to play a role in mediating PIT as their blockade have also resulted in attenuated or abolished transfer effects (Dickinson et al., 2000; Lex and Hauber, 2008). In tests observing DA receptor antagonists on the NAc core and shell on the performance of PIT, Lex and Hauber (2008) observed that infusions of D1 and D2 receptor antagonists on the core abolished PIT but these effects were less pronounced by D2 receptor antagonists. Similar effects were observed by D1 receptor antagonists in the shell. However, different doses of D2 receptor antagonists applied to the shell showed different effects with lower doses abolishing PIT and higher doses effective only during the initial presentation of the CS+. These results suggest that DA receptor activation are required for mediating the facilitatory effects of Pavlovian CSs on instrumental responding. In particular D1 receptor activation seem to have a more prominent role than D2 receptors in mediating these effects. However, the distinct effects observed by D2 receptor antagonists applied to the shell with respect to the core also suggests that D2 receptors might function differently in the shell and core. These studies summarised in tables 2.2 and 2.3 are among many which implicate the NAc and its DAergic influences in motivation and the acquisition of reward based behaviours. Other studies which implicate the NAc and its distinct subunits include cue induced goal directed behaviours (Floresco et al., 2008a; Yun et al., 2004; Fuchs et al., 2008).

Table 2.2: Experiments observing DAergic manipulations on the NAc

Reference	Task	Manipulation	Effect	Implications
Di Ciano et al. (2001)	Autoshaping	DA & Glu R antagonist on the Core	Disrupted acquisition and performance of discriminated approach behaviour	DA and Glu R involved in Pavlovian conditioned responding
Dalley et al. (2005)		DA & NMDA R antagonist on the NAc	Disrupted acquisition and performance of discriminated approach behaviour	D1 and NMDA receptors in the NAc essential for appetitive Pavlovian learning.
Parkinson et al. (2002)		DA depletion on the NAc	Impaired approach behaviour	DA release on the NAc mediates Pavlovian learning.
Dickinson et al. (2000)	PIT	DA antagonist on the NAc	attenuated PIT	DA release on the NAc is required for facilitating the effects of Pavlovian CSs on instrumental responding
Lex and Hauber (2008)	PIT	NAc D1 R antagonist Shell D2 R antagonist	High doses abolished transfer effect Delayed impairment in the transfer effect	DA D1 more than D2 R activation mediates the facilitatory effect of Pavlovian CSs on instrumental responding.
Yun et al. (2004)	Cue evoked goal directed behaviour	D1 R antagonism	Impaired responding to DS	NAc D1 R activity required for responding to cues that predict reward

Cue Evoked Goal Directed Behavior

In cue evoked goal directed behavioural experiments conducted by Yun et al. (2004), a discriminative stimulus (DS) was used to cue reward delivery contingent on a particular response. Inactivation of the NAc resulted in an increase in both rewarding and non rewarding responses. In addition, an increase in latency for the DS controlled response was also observed. In addition D1 receptor inactivation generated increased latency and attenuated responding to DS. These indicate that the NAc D1 receptor and NAc activation respectively are essential for and facilitate responding towards cues that signal rewards. In addition, NAc neurons are involved in exerting inhibitory influences on behaviours in response to the differential incentive values of cues with respect to their reward predictability (Nicola, 2007).

In assessing the specific roles of the NAc subregions in cue induced goal directed behaviours, Floresco et al. (2008a) observed that the NAc core and shell elicited rather distinct behaviours which implied that they functioned in opposition to one another. Core inactivated subjects demonstrated decreased reinstatement of extinguished responding (Floresco et al., 2008a; Fuchs et al., 2004; Chaudhri et al., 2008), while the opposite effect was observed in the shell lesioned subjects (Floresco et al., 2008a; Chaudhri et al., 2008). Interestingly, the application of GABA receptor agonists did not disrupt the ability of a reward predictive cue to influence Pavlovian approach, locomotory or instrumental behaviour (Fuchs et al., 2008; Floresco et al., 2008a). It was concluded by Floresco et al. (2008a) that the shell and core compete in opposing and complementary patterns for behavioural expression such that the core mediates the ability of CSs to influence instrumental behaviour, while the shell plays a role in updating the stimulus-reward contingency and facilitating alterations in behaviour depending on the current incentive value of the reward predicting stimulus.

These findings correlate with the experiments conducted by Corbit et al. (2001) which implicate the core in enabling actions that result in rewards and the shell in facilitating the effects of cues associated with rewards on

instrumental responding.

The NAc has also demonstrated involvement in feeding behaviour. Some studies are summarised in the following section and in table 2.4.

2.5.2 The NAc in Feeding

The NAc is positioned in such a way that it can be implicated in the control of feeding. The shell in particular receives information related to taste, visceral function and metabolic sensing such as the lateral hypothalamus (Kelley, 2004) and has been implicated in both feeding and the control of appetitive behaviour (Reynolds and Berridge, 2001). Maldonado-Irizarry et al. (1995) observed marked and prolonged feeding in satiated rats after Glu receptors in the shell and not the core, were blocked.

The distinct functionality of the NAc subregions in mediating rewarding and consummatory behaviours have further been demonstrated in experiments which observed the role of the NAc in feeding behaviours. In such experiments, shell manipulation by either excitatory glutamatergic AMPA receptor inactivation, inhibitory GABAergic receptor activation or its excitotoxic lesioning resulted in increased feeding behaviours or weight gain (Stratford and Kelley, 1997; Kelley, 1999a). With similar behaviour observed during LH stimulation, Maldonado-Irizarry et al. (1995); Kelley (1999a) demonstrated that the shell, through connectivity with the LH, influenced motivation and feeding behaviour (Kelley, 2004; Cardinal et al., 2002a). In addition, it was observed that the feeding elicited by the application of GABA agonists to the shell generated an increase in the synthesis of the nuclear protein Fos, in neurons in the LH (Stratford and Kelley, 1999). Fos expression confirmed interactions between these two regions.

While feeding behaviour has been affected by manipulations on the shell, experiments conducted by Reynolds and Berridge (2001) produced very interesting results whereby the application of GABA agonists on different regions of the shell produced distinct behaviours. GABA agonist application

Table 2.3: Experiments observing the NAc's role in instrumental and Pavlovian mechanisms

Reference	Task	Location	Effect	Implications
Blaiss and Janak (2008)	Goal-tracking Pavlovian Conditioned Approach (PCA)	core inactivation shell inactivation	decreased CS+ responding decreased CS+ responding increased CS- responding	Involved in process for expressing CS-US association Involved in the ability to ignore stimuli when rewards are unavailable
Parkinson et al. (2000)& Cardinal et al. (2002b)	Autoshaping	core lesion	decreased CS+ responding	Site for which Pavlovian CSs mediate instrumental responding
Yun et al. (2004)	Cue evoked goal directed behaviour	NAc inactivation	Increased responding to of specific behaviours	Involved in facilitating and suppressing responses to cues with different incentive values
Corbit et al. (2001)	Outcome devaluation Devaluation extinction	core lesion core lesion	General decrease in responding failed to show selective effect of the devaluation treatment	implicated in instrumental performance mediates action selection based on the value of its outcome
Floresco et al. (2008a)	Cue induced reinstate-ment	core inactivation shell inactivation	Attenuated response to reinstated CS Enhanced response to reinstated CS	Mediates CS induced instrumental responding Enables the change in incentive value of cues to mediate responding

on the rostral region of the shell resulted in increased feeding behaviour. On the other hand, GABA agonist application on the caudal regions of the shell elicited defensive burying behaviours. These results demonstrate how the shell as a unique heterogenous structure receives information from both the external environment and internal regions (Kelley, 1999b) and functions as a unit capable of controlling distinct behaviours. Such ability to access different characteristic functions by manipulating different shell regions proposes that the shell might also play a role in mediating the adjustment of basic behaviours dependent on reward predictability. This characteristic is useful for mediating behavioural flexibility which occur in different forms including set-shifting, and reversal learning both of which are discussed next.

2.5.3 The NAc in Spatial Learning and Behavioral Flexibility

Reversal learning is a simpler form of behavioural flexibility whereby a discrimination is required between two or more stimuli such that only one of the stimuli is associated with a reward. Once this discrimination has been learned, a shift occurs and the reward is omitted from the stimulus that was initially associated with it and associated with another stimulus. Behavioral flexibility involves an ability to adapt according to this change. Set-shifting is a more complex form of behavioural flexibility that requires shifts between strategies, rules or attentional sets (Floresco et al., 2008b).

The performance of rats in spatial discrimination, set-shifting and reversal learning have been studied by many (Birrell and Brown, 2000; Egerton et al., 2005; Floresco et al., 2006, 2008b). In such experiments rats with NAc lesions demonstrated impairment in both the learning and reversal of a T-maze spatial discrimination task and impairment in the learning of a hidden platform in a water maze. The lesioned rats required more trials to meet the criterion and made more errors until criterion was met. In particular, lesioned rats were slower at relearning the location of the reward (Annett et al., 1989). While other studies involving NAc lesions did not reveal impairments in

Table 2.4: Experiments observing the role of the NAc in feeding and behavioural flexibility

Reference	Task	Location	Effects	Implications
Reynolds and Berridge (2001)	Feeding and fear	GABA R Agonist on rostral shell GABA R Agonist on caudal shell	Increased appetitive feeding behaviour Elicit defensive behaviour	Shell mediates feeding and defensive behaviours
Stratford and Kelley (1997) Stratford and Kelley (1999)	Feeding	GABA Agonist on the shell	Increased feeding behaviour Fos expression in the LH	Shell is involved in feeding behaviour Shell-LH link involved in feeding behaviour
Fuchs et al. (2004)	Cue induced reinstatement	core inactivation	Abolished reinstatement	Core essential for cue induced reinstatement
Annett et al. (1989) & Stern and Passingham (1995)	Reversal learning	NAc lesion	Impaired spatial reversal learning	NAc implicated in reversal learning
Floresco et al. (2006)	Set shifting	Core inactivation Shell inactivation	Disrupted acquisition & maintenance of strategy Improved performance during set shift	Facilitates acquisition of strategies Involved in inhibiting behaviour

initial discriminations or particular reversals. Combined, these studies led Floresco et al. (2008b) to imply that the NAc might be involved in specific reversals.

The facilitatory role DA plays in mediating spatial learning and behavioural flexibility were further confirmed by the observed effects of DA depletion in the NAc. These effects included impairments in the acquisition of spatial discrimination, in the ability to alternate behaviours, as well as to reverse previously acquired behaviours (Taghzouti et al., 1985a). In addition, to DA manipulation on the NAc, ibotenic lesions of the NAc have also resulted in impaired spatial reversal (Stern and Passingham, 1995).

The dissociable role of the NAc subregions were analysed by observing the performance of core and shell lesioned rats in strategy set shifting experiments (Floresco et al., 2006). Here, rats had to shift between visual cues and response strategies. The errors made during the experiments were scored as follows: Errors were scored as perserverative when the rats made incorrect choices towards the previously rewarded strategy 75% of the time. When these errors were made 50% of the time or less, they were scored as regressive errors. Errors were classed as never-reinforced when the rats made errors during both the initial discrimination and during the initial strategy shift. Inactivation of the core immediately before strategy set shift required more trials to reach criterion and resulted in more regressive errors. Therefore, core inactivation impairs the ability to acquire and maintain new strategies. Shell inactivations made prior to the initial discrimination training rather than prior to the strategy shift showed a reduction of the required number of trials to reach criterion. These findings summarised in table 2.4, suggest that the shell is involved in mediating learning to inhibit responding to non-rewarding cues and led Floresco et al. (2006, 2008b) to conclude that the NAc shell and core subregions are involved in dissociable roles and they respectively mediate learning about irrelevant stimuli and the acquisition as well as maintenance of new strategies.

2.5.4 The NAc in Latent Inhibition

Latent inhibition (LI) occurs when the learning of conditioned associations to a stimulus is retarded due to prior exposure to the stimulus. An example of LI can be observed by comparing two groups of rats in experiments in which a CS is paired with a US. Prior to this pairing the first group of rats were pre-exposed (PE) to the CS but no US was initially paired with the CS. In the second group, the rats were not pre-exposed (NPE) to any CS or US. It was observed that the rats from the PE demonstrated reduced acquisition of the CS-US association than the NPE group. The PE group showed LI. Disrupted LI has been proposed as a model for schizophrenia. Both the mesolimbic DA projections and their release on the NAc have been involved in the control of normal, disrupted and potentiated LI (Solomon and Staton, 1982; Weiner, 2003). The systemic application of DA releasing nicotine or low doses of amphetamine showed disrupted LI (Weiner et al., 1987; Weiner, 2003; Joseph et al., 2000). In addition, increased and decreased DAergic activity on the NAc generated enhanced switching and perserverative behaviours respectively (Weiner, 2003; Taghzouti et al., 1985b). In particular, potentiated LI was observed in animals with either DA depletion in the NAc or application of D2 receptor antagonists (Joseph et al., 2000). Furthermore, just as the shell and core lesions have been observed to differentially affect set shifting, so also have such lesions disrupted LI in different ways. While LI was persistent under conditions that disrupt LI or left intact in core lesioned agents, shell lesions demonstrated disrupted LI. These findings suggest that the shell and core generate opposite effects such that the shell is involved in modulating and inhibiting switching mechanisms while the core simply enables responding according to stimulus reward contingencies (Weiner, 2003).

Although shell lesions do not impair Pavlovian approach behaviour or instrumental conditioning (Parkinson et al., 1999, 2000), the shell seems to facilitate the invigorating effects of rewards on behavioural responses (Ito et al., 2004). Lesion studies done by Corbit et al. (2001) also suggest that the shell plays a role in transferring associations obtained between stimulus

and rewards on to instrumental responding. In addition, inactivation of different regions of the shell have been implicated in eliciting distinct appetitive and defensive behaviours (Reynolds and Berridge, 2001). Therefore while the core enables motor activity towards reward predicting stimuli, the shell facilitates alterations in behaviour when a change in the incentive value of the reward predicting stimulus occurs (Floresco et al., 2008a). Based on inactivation and lesion experiments, the core enables all reward related behaviours to be driven by their associated stimuli and the shell seems to play an essential role in enabling behaviour with the highest probability of obtaining rewards to dominate and adjust when the incentive value of the stimulus predicting the reward changes.

With the shell and core differentially implicated in a variety of reward based behaviours and DA transmission on these regions playing a major role in reward based learning, the distinct mechanisms implemented by the shell and core sub units might involve unique methods by which DA is released on the sub-units. The following section addresses a study which suggests that DA is transmitted differentially in the shell and core.

2.5.5 Differential DA transmission on the NAc

DA transmission in the NAc core and shell have been established to play an important role in behaviour motivated by reinforcers. With a variety of studies observing different behavioural effects when manipulating the NAc subregions, Bassareo and Di Chiara (1999); Bassareo et al. (2007) conducted certain studies to confirm the change in DA transmission in the shell and core during appetitive and consummatory phases of behaviour motivated by food (Bassareo and Di Chiara, 1999) and drugs (Bassareo et al., 2007). It was observed that drug CS resulted in potentiated DA release in the shell and not in the core. In contrast, food CS phasically stimulated DA transmission in the core rather than in the shell. These studies along with confirming that the shell and core differentially mediate reward based behaviours, also show that DA transmission is released at different levels in the shell and in the

core.

2.6 Concluding Remarks

In this chapter, the NAc and its shell and core sub units have been introduced as an input structure to the basal ganglia. The NAc functions as an important interface through which the motivational effects of reward predicting cues and stimuli obtained from limbic and cortical regions transfer onto response mechanisms and instrumental behaviours (Di Ciano et al., 2001; Cardinal et al., 2002a,b; Balleine and Killcross, 1994). The NAc and the DA neurons of the VTA are innervated by excitatory glutamatergic neurons of limbic and cortical regions as well as the lateral hypothalamus which can be activated by primary rewards.

Reward predictive cues have been observed to excite regions of the NAc (Nicola et al., 2004) which when lesioned have demonstrated a reduction of the rewarding effects of drugs (Roberts et al., 1977) and instrumental responding (Balleine and Killcross, 1994).

Some studies which focus on manipulating the NAc connectivity so as to obtain improved understanding of its functionality with respect to motivation and reward based behaviours have been summarised. The functionality of the NAc shell and core subregions as distinct sub units and the contribution of certain afferent structures on behaviour have been provided. Overall the studies summarised here are among many which support the role of the NAc in controlling response selection by integrating a range of information from a variety of input regions.

The NAc receives projections from the cortex to form corticostriatal connections. In addition, the NAc is also innervated by dopaminergic neurons. The corticostriatal connections can undergo both LTP and LTD mediated by DA receptor activation. In particular, D1 receptor activation mediates the induction of both LTP and LTD while D2 receptor activation seems to favour the

induction of LTD (Yin et al., 2006; Reynolds and Wickens, 2002; Calabresi et al., 1992c; Kerr and Wickens, 2001; Calabresi et al., 2007). The production of LTP and LTD are also influenced by pre- and postsynaptic activities. There are two pathways that have been summarised which may influence the different concentration levels of dopamine in corticostriatal synapses. A phasic burst in DA activity in event of rewards results in DA release at concentration levels that could activate D1 receptors (Goto et al., 2007). On the other hand tonically active DAergic neurons produce DA at concentration levels that might favour D2 receptor activation (Goto and Grace, 2005b).

While the core enables reward related behaviours to be driven by their associated stimuli, the shell seems to play an essential role in enabling behaviour with the highest probability of a reward to dominate and adjust when the incentive value of the stimulus predicting the value changes. The shell also seems to play a role in mediating switching in basic behaviours by inhibiting irrelevant responses. The shell can influence activity on the DA neurons of the VTA through a direct inhibitory shell-VTA pathway and through an inhibitory shell-VP-VTA pathway (Zahm and Heimer, 1990). The shell can also influence core activity via an indirect shell-VP-MD-cortical-core pathway (Zahm and Heimer, 1990; Groenewegen et al., 1999; Zahm, 2000). In addition to the distinct connectivity between the shell and the core, the shell's ability to influence both DA and core activity might be useful in describing how uniquely these subunits function. Differential DA transmission in the core and shell have also been observed which could further be used to explain how the shell and core function differently.

The functionalities, processes and mechanisms involved as well as the underlying assumptions made regarding the behaviour and contribution of the NAc core and shell circuitry are summarised in table 2.5. These will be considered when developing the computational model in the following chapter. These distinct roles of the NAc shell and core subunits have been documented so that a model circuitry can be developed accordingly in the following chapter.

Table 2.5: The established functionalities and assumptions based on the biological constraints

Region/Effect	Functionality & Characteristics	Assumptions
1. Medial prefrontal cortex (mPFC)	Sensor inputs to the core.	
2. Orbitofrontal cortex (OFC)	Sensor inputs to the shell.	
3. Lateral hypothalamus (LH)	Innervates the VTA and shell.	Influences phasic DA activity
4. Shell	Mediates facilitation of highly rewarding behaviours to dominate and facilitates changes in behaviour.	Strong LTP and LTD
	Connected to the DA VTA neurons via the shell-VP-VTA pathway.	Influences tonic DA activity
5. Mediodorsal nucleus of the thalamus (MD)	Connectivity between shell and core.	Shell facilitates and attenuates core activity
6. Core	Enables reward driven motor behaviour	Strong LTP and weak LTD
7. Phasic DA activity	Activates DA D1 receptors which mediates corticostriatal LTP.	D1 R activation induces LTP
8. Tonic DA activity	Activates DA D1 and D2 receptors which mediate corticostriatal LTD.	D2 R activation induces LTD

Chapter 3

Developing a Computational Model of the Nucleus Accumbens Circuitry

3.1 Introduction

In the previous chapter, the circuitry surrounding the nucleus accumbens (NAc), the plasticity occurring in this region and its overall role in a variety of reward based behaviours have been described.

A computational model has been developed based on the experimental observations and assumptions made in the previous chapter. It has been integrated into an agent and used to learn, adapt and demonstrate reward seeking behaviours similar to those obtained in biological organisms. The afferent structures to the NAc modelled as a generalised input system, are specialised when required according to each of its unique characteristics. As such the overall system will be divided into an input, reward and adaptive system. The adaptive system based on the limbic circuitry has been developed so that it is sufficient in performing reward based learning. It will be shown how this system, analogous to the NAc, can be further specialised into

the shell and core subunits.

In chapter 2, it was concluded that the shell plays an essential role in enabling behaviour with the highest probability of a reward to dominate and adjust when the incentive value of the stimulus predicting the value changes. The shell also seems to mediate the switching between basic behaviours by inhibiting irrelevant responses (Floresco et al., 2008a). The core on the other hand seems to enable behaviour in response to the reward predicting stimuli (Floresco et al., 2008a). The adaptive systems sub units will be modelled to possess the characteristics of the shell and the core as described.

A description of how this biologically inspired model surrounding the NAc circuitry has been developed and integrated into an agent which utilises signals from an environment is provided. The agent interacts with the environment and learns to complete reward seeking tasks. In this chapter, the concept is demonstrated in open-loop experiments during which it will be shown to account for some basic features of classical conditioning. In this way, the models performance in the open-loop experiments can be compared against empirical studies of classical conditioning.

A range of classical phenomena including spontaneous recovery, re-exposure to the reinforcer (Rescorla, 1972; Rescorla and Heth, 1975; Bouton, 1984) and savings in reacquisition (Napier et al., 1992), support the theory that the originally learned associations stay preserved during extinction and that learned associations that enable behaviour are suppressed rather than unlearned. In this chapter the model will be developed in which associations are learned between cues and rewards. When these cues no longer precede reward, the model adapts without extinguishing the learned associations that generate responding.

In the following chapter, this model will be implemented so that it can perform in a closed-loop biologically inspired behavioural tasks. Its capability will be tested in reward seeking scenario experiments.

3.2 The Environment and the Agent

In this section a learning agent is introduced. The agent exists in an environment (Fig. 3.1) in which it finds rewards and learns to predict such rewards using signals from the environment that correlate temporally with reward delivery. The reward (r) signals a biologically significant stimulus and occurs during the unconditioned stimulus (US). When the agent receives an US, it elicits a response referred to as the unconditioned response (UR).

The signals that precede the US are initially neutral. Each neutral signal will be referred to as a conditioned stimulus (CS) because of their individual potential to become associated with the US. The agent acquires an association between these CS and the US so that they are referred to as the conditioned stimulus (CS). The CS inputs are capable of eliciting a learned response which generally precede US and reward delivery and indicates that the agent has learned to predict the availability of a reward.

The agent embedded in the environment is shown in Fig 3.1. The signals generated by the environment comprises a reward and a number n of conditioned stimuli. The agent generates a set of behaviours which initially occur in response to the US and can be learned in response to the CS. As such the CS inputs can also be referred to as predictive inputs.

The agent in Fig. 3.1 comprises an input system, reward system, and an adaptive system which integrates and adapts to information from the input and reward systems. The adaptive system is capable of enabling the agents actions in response to the CS and the US. The adaptive system is made up of two learner units namely, the shell and the core units. The role of the core sub unit is to learn to enable actions that lead to rewards depending on the information processed at the inputs. As the environment changes, the inputs which once predicted reward delivery can change such that no more rewards are made available. The stimuli which originally predicted the rewards become irrelevant and their incentive value is reduced. The agent must adapt to these changes and learn to respond to inputs accordingly. The

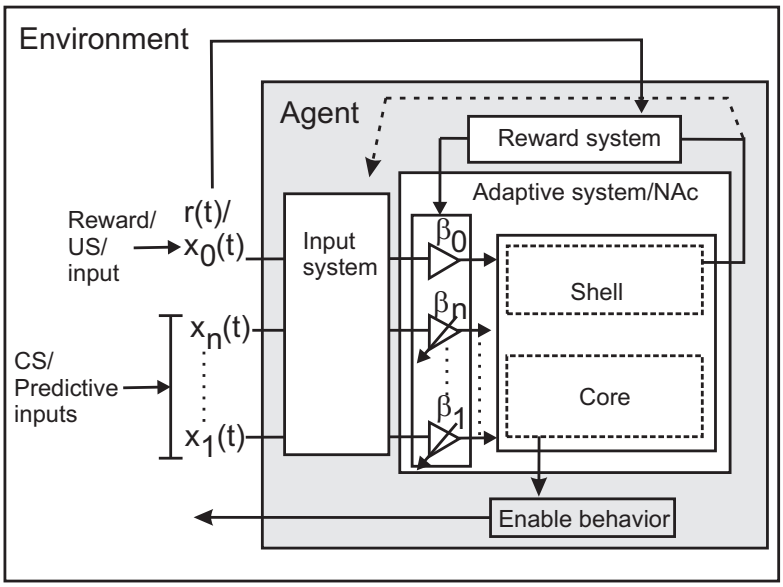

Figure 3.1: The agent integrated into the environment comprising of the input, reward and adaptive systems. The environment generates the reward (r) or US and the CS. These are processed by the agents input and reward systems. The adaptive system is capable of enabling the agents actions to generate behaviour in the environment.

shell unit learns which inputs are "*valuable*" to the agent and updates the adaptive system accordingly.

The reward system of the agent processes information based on whether or not the agent obtains a reward in the environment. This reward system can be modelled as the dopaminergic (DAergic) neurons of the ventral tegmental area (VTA) characterised by its two dopamine (DA) transmission modes.

3.3 Modeling the Reward System

The reward system is modelled by the VTA DA neuron's spiking activities which is influenced by its afferent structures (table 2.5(3)). One of the VTA's afferent units, which becomes active when a primary reward (such as food) is obtained, is the lateral hypothalamus (LH) (Kelley et al., 2005). The LH

is represented by the filtered reward signal r(t).

$$LH(t) = r(t) * h_{LP}(t) \tag{3.1}$$

The temporal convolution $*$, is a technique implemented in signal and image processing (Croft et al., 2001). It is used to calculate the output of the nuclei representation. This output can be obtained by convolving the input signal $r(t)$ with the lowpass filter (h_{LP}) which represents the LH.

Figure 3.2A shows how the lowpass filter (h_{LP}) representation of LH transforms a δ-pulse or reward signal input r into a damped oscillation (Porr and Wörgötter, 2001; Porr and Wörgötter, 2003). Lowpass filters or resonators are used to simulate the biophysical characteristics of neuronal signal transmission (Porr, 2004; Shepherd, 1998). For example, an action potential can be represented by the impulse response of the lowpass filter. The lowpass filter is defined according to the Laplace transform as follows:

$$H_{LP}(s) = \frac{K}{(s+p_1)+(s+p_2)} \tag{3.2}$$

K is a constant and the poles are defined:

$$p_1 = a + jb \tag{3.3}$$

$$p_2 = a - jb \tag{3.4}$$

where the real (a) and imaginary (b) parts are defined by:

$$a = \frac{\pi f}{q} \tag{3.5}$$

$$b = \sqrt{(2\pi f^2) - (\frac{\pi f}{q})^2} \tag{3.6}$$

The frequency of oscillation and damping characteristic or quality factor of the filter are given by f and q respectively.

The lowpass filter is also used to extract signals which occur at certain low

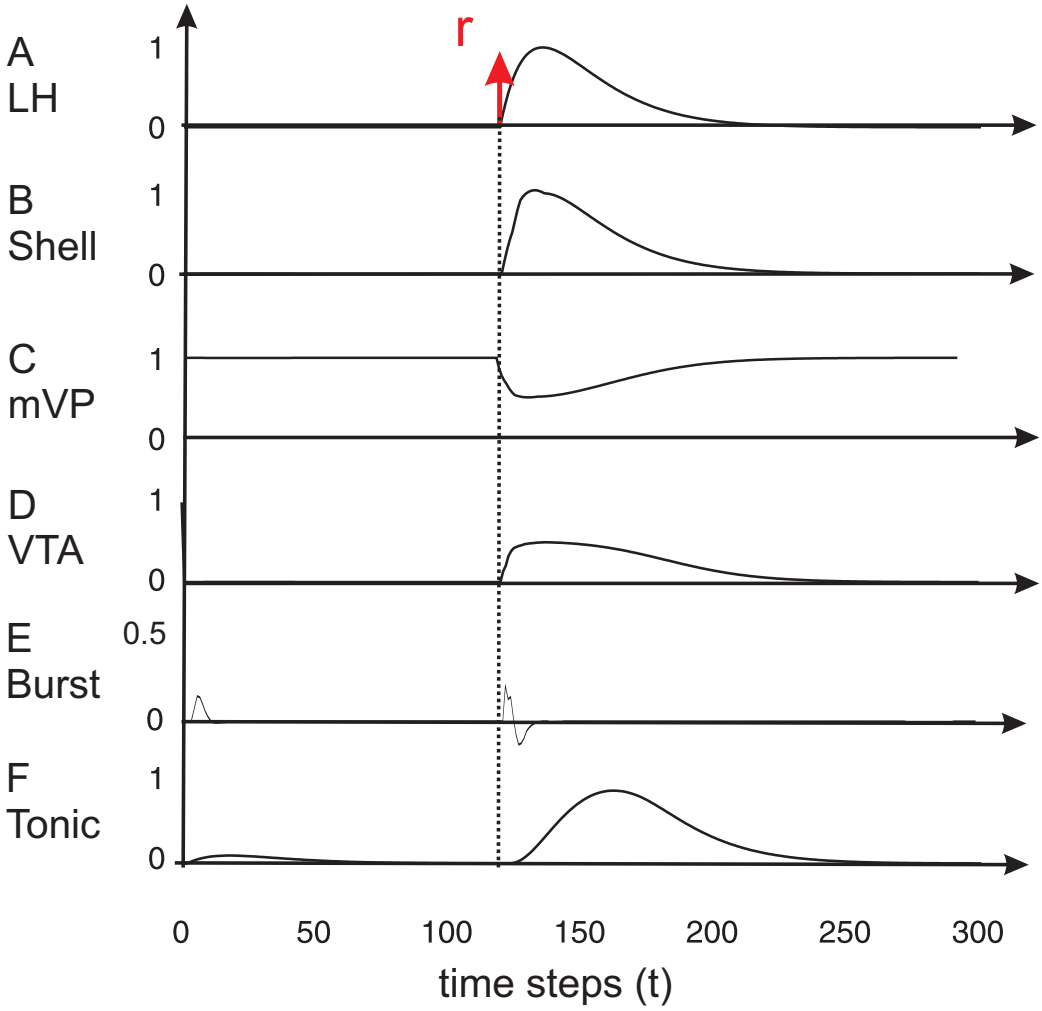

Figure 3.2: A) The LH represented by a lowpass filter (h_{LP}) transforms the reward signal, a delta pulse, into a damped oscillation. B) The shell activated by the LH projection. C) The mVP activity is inhibited by the shell. D) The VTA is dis-inhibited by the shell via the mVP inhibition. E) The burst, represented by a passing the VTA signal through a highpass filter (h_{HP}). F) The tonic activity, represented by passing the VTA signal through a lowpass filter. The signals observed at the burst and tonic panel between the time steps 0 to 50 are simulation startup artefacts. Parameters: The h_{LP} frequency and q-factors are set to 0.01 and 0.51 respectively. The h_{HP} frequency and q-factors are set to 0.1 and 0.71 respectively. The learning rate = 0.6. $\kappa = 1$, $\upsilon = 1$ and $\eta = 1$. $\theta_{tonic} = 0$ and $\theta_{burst} = 0.05$. $\chi_{burst} = 1$ and $\chi_{tonic} = 0.1$.

frequencies. Signals occuring at higher frequencies can in turn be detected by a highpass filter which is represented in Eq.3.8. All filters implemented in this model are identical to the resonators implemented in Porr and Wörgötter (2001); Porr and Wörgötter (2003).

The VTA is innervated by the medial ventral pallidum (mVP) and shell nuclei both of which also contribute to the VTA's spiking activity. The individual mVP and shell release inhibitory GABAergic neurotransmitters. The mVP is also inhibited by the shell and in turn actively inhibits the VTA. When the shell is stimulated, the mVP becomes inhibited and its active inhibition on the VTA is reduced. The shell can be identified as an adaptive system. The shell and mVP discussed later on in this chapter, are respectively represented in equations 3.19 and 3.20. DA neurons of the VTA innervated by the LH, mVP and shell are summarised:

$$VTA(t) = \frac{1 + \kappa \cdot LH(t)}{1 + \upsilon \cdot mVP(t) + \eta \cdot Shell(t)} \tag{3.7}$$

Here the VTA is activated by the LH and inhibited by the mVP and shell activities. υ and η are constants which represent the connecting weights of the mVP and shell to the VTA respectively. Figure 3.2B, C and D illustrate the shell, mVP and VTA activity respectively. The shell, activated by the LH, inhibits the mVP which actively inhbits the VTA. Therefore the increased activity in the shell results in an activation of the VTA. The connectivity between the LH, VTA shell and mVP is illustrated in Fig.3.3.

There are two spiking activities of VTA DA neurons. When the reward has been obtained, the VTA fires in high frequency burst spikes (Grace, 1991; Floresco et al., 2006). These high frequency bursts can be detected by passing the VTA through a highpass filter. The highpass filter is defined according to the Laplace transform:

$$H_{HP}(s) = \frac{s^2}{(s + p_1) + (s + p_2)} \tag{3.8}$$

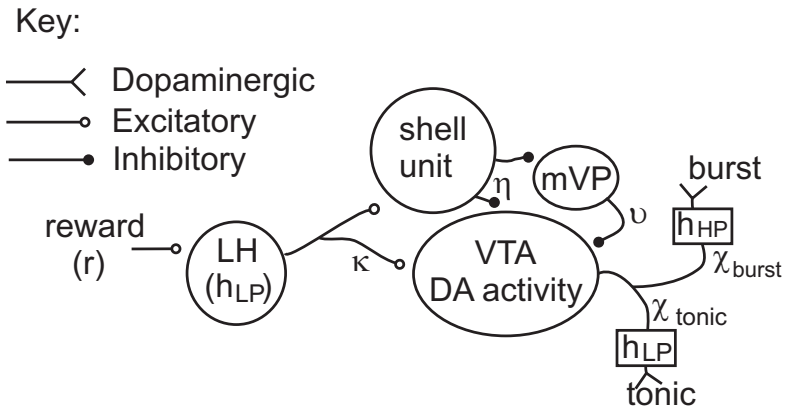

Figure 3.3: A representation of the connectivity between the LH, shell, mVP and VTA. Tonic and burst DA activities are generated by passing the VTA through a lowpass and highpass filter respectively.

p_1 and p_2 are the poles defined in Eq.3.1 and Eq.3.4.

The highpass filter responds to the high frequency bursts of the VTA:

$$burst(t) = \begin{cases} 1 & \text{if } [\chi_{burst} \cdot\ VTA(t) * h_{HP}(t)] \geq \theta_{burst}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

χ_{burst} represents the fraction of VTA neurons which are passed through the highpass filter. The highpass filter which represents the burst is illustrated in Fig.3.2E and in Fig.3.3.

The burst equates to 1 when the high frequency component of the VTA reaches a certain threshold value θ_{burst}. The threshold value represents the dopamine D1 receptors that become activated in event of DA bursts. D1 receptor activation plays a major role in mediating reward based learning (table 2.5(7)). The second activity of the VTA occurs in the absence of rewards when the population of its tonically active neurons increase. Such increase in the population of tonically active neurons are influenced by the inhibition of the inhibitory GABAergic neurons of the mVP which target the VTA. The mVP is also inhibited by GABAergic neurons of the NAc shell.

This means that the resultant effect of the shell on the VTA via the mVP is dis-inhibitory. DA released in this manner occurs over a very slow time course (Grace, 1991; Floresco et al., 2006) and is represented here by passing the VTA neurons weighted by χ_{tonic} through a lowpass filter as shown in Fig.3.2F and in Fig.3.3.

$$tonic(t) = \Theta_{tonic}(\chi_{tonic} \cdot [VTA(t) * h_{LP}(t)]) \tag{3.10}$$

where

$$\Theta_x(y) = \begin{cases} y & \text{if } y > \theta_x, \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

This tonic DA activity can be used by the agent to decode when rewards are unavailable and thus a flexibility in behaviour can be achieved when contingencies change in the environment (table 2.5(8)). With the mechanisms for encoding if and when rewards are obtained formalised, the next section describes a basic method by which the input system processes the signals obtained from the environment. This can be specialised to model the characteristics of the afferent structures to the NAc according to their ability to influence behaviour mediated through the NAc.

3.4 Modeling the Cortical Input System

The signals from the environment which represent the CS and US are processed by the agent's input system (Fig. 3.1) and are indexed x_0 for the US input and x_j where $0 < j \leq n$ for $\mathbf{n}$ predictive inputs.

The US or predictive inputs can trigger the agents motor-enable system such that the agent always elicits behaviour in response to the stimulus. All the input signals generate internal representations of the stimuli which project onto the adaptive system. Just as the reward is passed through neuronal resonators, the internal representation of the CS and US are also generated by passing the signals through a resonator.

$$u_0(t) = h_{LP}(t) * x_0(t) \tag{3.12}$$

This filtered US input (u_0) projects onto the adaptive system through a fixed weighted channel. The adaptive system has access to enabling the agents behavioural responses. Therefore, the US input is capable of directly activating the agents motor action to generate the UR. The predictive inputs also feed into the adaptive system. Along with the US input, they correspond to the stimuli which can be processed by the prefrontal cortex (PFC). The PFC acquires and internally maintains information from recent sensory inputs to enable goal directed actions (Funahashi et al., 1989; Durstewitz and Seamans, 2002). This ability exhibited by the PFC is known as working memory, whereby earlier stimulus are capable of elevating and retaining activity over delay periods. The PFC maintains persistent activity triggered from cues from the environment for a set period or until a primary reward is obtained. Similar to the US pathway, the **n** predictive inputs can be filtered:

$$u_{pre-j}(t) = h_{LP}(t) * x_j(t) \tag{3.13}$$

where $0 < j \leq n$. In addition, their ability to undergo persistent activity is illustrated in Fig. 3.4 and represented as follows:

$$u_j(t) = \begin{cases} 0 & \text{if } LH_{reset}, \\ PA_{max} & \text{for period } T_{PA}, \\ & \text{if } u_{pre-j}(t) \geq PA_{max} \\ u_{pre-j}(t)(t) & \text{otherwise.} \end{cases} \tag{3.14}$$

Cortical neurons become activated prior to reward presentation and terminate after the reward has been delivered (Schultz et al., 2000). T_{PA} represents a set duration by which the predictive channel maintains activity when the activities reach a threshold value (PA_{max}). Since the primary reward is signalled by the LH activation, the activity in these neurons are suppressed with

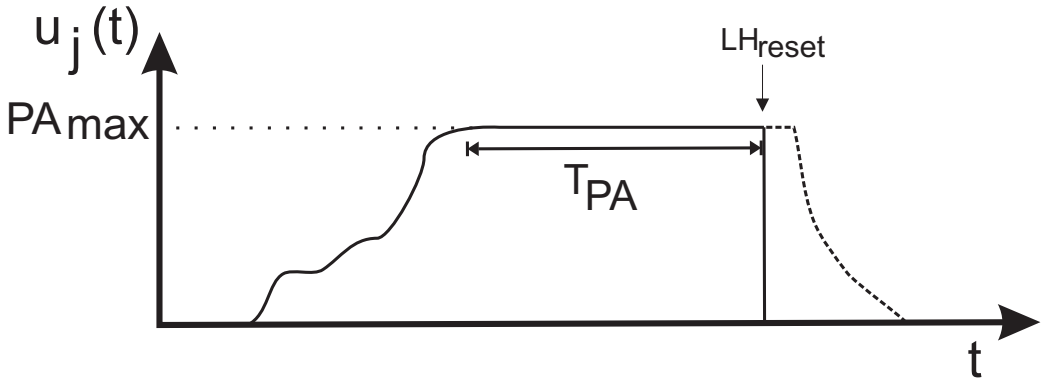

Figure 3.4: Persistent activity occurs for a duration of T_{PA} when $u_j(t)$ reaches a value PA_{max} or until the LH input which signals the reward resets the activity.

respect to the LH activity (LH_{reset}).

The predictive pathways are capable of undergoing acquisition and generating learned or conditioned responses. These predictive x_j signals are filtered (u_{pre-j}) and fed into the adaptive system through plastic weighted channels β_j. The plastic weights change according to a differential Hebbian learning rule (Roberts, 1999; Porr, 2004). These plastic weights, modified in the learning process, alter future behavior. In the model the input system will be identified as cortical units which project to the NAc however , the inputs can also be specialised to model limbic inputs that provide emotional, spatial or contextual information to the NAc (Cardinal et al., 2002a). The filtered predictive inputs u_j which simulate the cortical inputs and innervate the individual core and shell units of the adaptive system will be specialised as the prefrontal cortex (PFC) and orbitofrontal cortex (OFC). The PFC and OFC are used to indicate which cortical units respectively innervate the core and shell (table 2.5(1 and 2)). The plastic weights of the predictive channels are susceptible to change and are therefore capable of enabling behaviour in response to the predictive inputs. Weight change and the adaptive system are described next.

3.5 Modeling the NAc Adaptive System

The input and reward systems were represented in the previous section with the predictive inputs transmitting information via weighted channels. The weights change in the adaptive system and will initially be described using a general term β. The adaptive system will be specialised into its shell and core subunits. Accordingly, the general term β will be specialised to ω and ρ weights of the shell and core sub units respectively.

The filtered input signals sum onto the adaptive system (Fig 3.1) through weighted channels (β) as follows:

$$V(t) = u_0(t) \cdot \beta_0(t) + \sum_{j=1}^{n} u_j(t) \cdot \beta_j(t) \tag{3.15}$$

The US component (u_0) feeds into the adaptive system through a fixed weight (β_0) (Porr and Wörgötter, 2003) while the predictive $u_j(t)$ filtered inputs transmit information to the NAc through their respective plastic weights (β_j). In a naïve agent, these plastic weights are set at zero and change as learning occurs. Eventually as the adaptive system learns, the plastic weights change such that the change in the output V(t) is proportional to the sum of the US inputs and the predictive inputs that suitably predict the reward. The plastic weights adapt and undergo long term potentiation (LTP) or long term depression (LTD) depending on whether they increase or decrease respectively. Weight change is described and modelled in the next section.

3.5.1 Weight Increase: Isotropic Sequence Order Learning and the Third Factor (ISO-3)

The plastic weights in the adaptive system undergo change as illustrated in Fig. 3.5. For simplicity, the adaptive system is analysed to process two δ-function input signals. The US input signal (x_0) and a predictive signal

(x_1) obtained at a temporal event before the US input (Fig. 3.5B1). This does not mean that the adaptive system's capability is limited to processing only two input signals. The output $V(t)$ is an accumulation of the weighted predictive and US inputs:

$$V(t) = \beta_0 \cdot u_0(t) + \beta_1 \cdot u_1(t) \tag{3.16}$$

The weight of the US input β_0 is fixed, while weight change for the predictive input β_1 determined by a learning rule known as three factor isotropic sequence order (ISO-3) learning (Porr and Wörgötter, 2003). This is isotropic sequence order learning (ISO) (Porr and Wörgötter, 2003; Porr, 2004) that is enabled by a third factor. ISO learning is a form of differential Hebbian learning in which learning occurs depending on the temporal correlation between two neuronal activities. ISO-3 learning is defined as follows:

$$\Delta\beta_1(t) = \mu \cdot u_1(t) \cdot V'(t) \cdot burst(t) \cdot (limit - \beta_1(t)) \tag{3.17}$$

The rate of weight change is set according to the learning rate μ. Here $u_1(t)$ and $V'(t)$ represent the pre-synaptic and the derivative of the post-synaptic activities respectively (ISO Learning). The burst represents a third factor and turns ISO learning into ISO-3 learning. In this work the burst is a reward signal represented by phasic dopaminergic. The weight change is bound such that the maximum value it can reach is determined by the "*limit*". This form of weight change can also be described as three factor correlation based differential Hebbian learning (Porr and Wörgötter, 2003).

This form of learning implemented in the agent is capable of changing any number of weights whose inputs precede the US input within a certain time frame. This means that an 'association' develops for any input that suitably precedes any US input in an agent required to learn to predict from a multiple number of predictive inputs. The burst triggers learning only during relevant moments (i.e. an association can be formed between the US input that enables behaviour which results only in the delivery of a reward). Using a

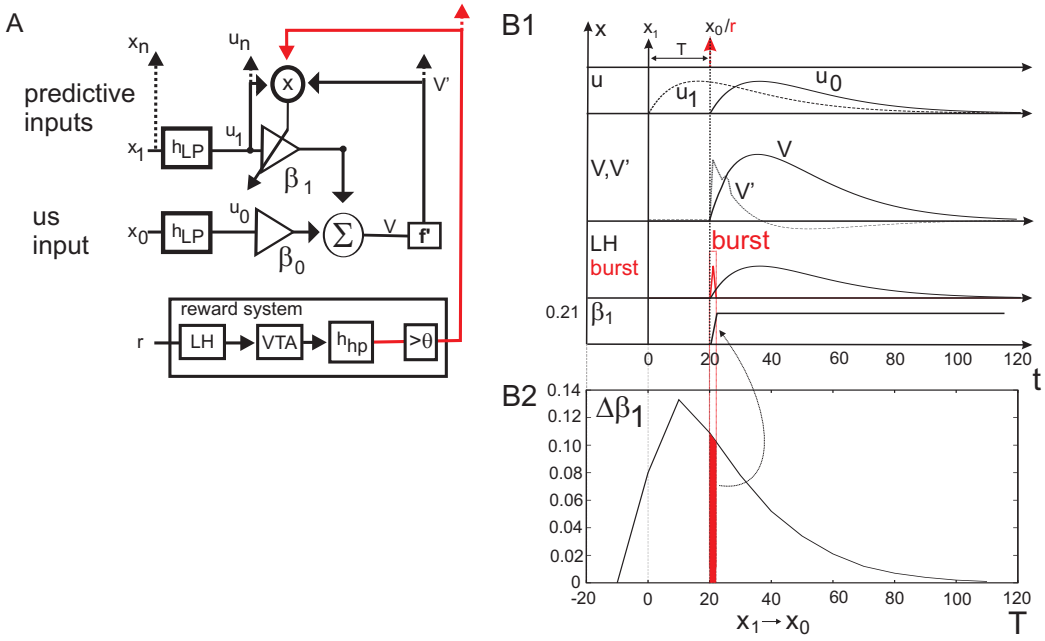

Figure 3.5: A) The circuit diagram representation of ISO-3 learning. The n CS inputs x_1 to x_n are filtered to become u_1 to u_n respectively. The US input x_0 becomes filtered to generate u_0. The CS inputs influence the output signal V through plastic weights each of which change according to the correlation between the corresponding filtered CS input signal, the output derivative V' and a third factor. The output V is a combined sum of the CS and US inputs. The 3rd factor which enables the weight β_1 (initially set at 0) to change is generated by the reward system. B1) ISO-3 learning signal traces when two δ pulses representing the predictive and US inputs which are filtered and correlate during the burst to generate a weight increase on the plastic predictive weight β_1. B2) The weight change curve as a function of T, the duration between the CS and US input. The change of β_1 is dependent on the duration of the burst. Therefore, the weight development during this correlation, is represented by the shaded area under the curve indicated by the arrow. Parameters: The h_{LP} frequency and q-factors are set to 0.01 and 0.51 respectively. The h_{HP} frequency and q-factors are set to 0.1 and 0.71 respectively. $\beta_0 = 1; \theta = 0.025$. (Thompson et al., 2006)

third factor as a gate has the advantage of minimising the destabilising effects that may drive learning outside the event of the gating signal (Thompson et al., 2006; Porr and Wörgötter, 2007).

The third factor can be represented by DA burst spiking activity. DA bursts occur when primary rewards are obtained and activate the LH which excite the DA neurons of the VTA to fire (Eq. 3.7 and Eq. 3.9).

The circuit diagram illustrating ISO-3 learning is shown in Fig. 3.5A. The signal traces occuring between a CS and US input during learning is illustrated in Fig. 3.5B1. The weight change is dependent on the filtered CS input, the output derivative and the third factor. It can be seen how the third factor DA burst enables learning only when it is triggered. The weight change curve for $\beta > 0$ when $T \geq 0$ is shown in Fig. 3.5B2. The weight change is proportional to the duration of the DA burst and increases according to the shaded area under the curve. The magnitude by which the weight changes is very much dependent on the timing (T) between the predictive and US inputs relative to the US input (Porr and Wörgötter, 2003; Porr, 2004) as well as the duration of the burst. Therefore, the weights in the adaptive system are capable of increasing depending on their pre and post-synaptic activities and DA burst. ISO-3 learning is analogous to LTP generated in event of pre-synaptic release of glutamate, post-synaptic depolarisation and DA D1-type receptor activated by the higher DA concentration levels generated by bursting DA transmission.

While the DA burst enables the predictive weights to increase at significant moments with respect to the ISO learning rule, the second DA activity is modelled to enable weight decrease on previously learned plastic synapses in the absence of relevant moments. The following section describes how learned plastic weights in the adaptive system are capable of decreasing depending on the tonic DA activity.

3.5.2 Weight Decrease (LTD)

The plastic weights in the adaptive system decrease (LTD) depending also on DA activity. The second spiking activity of DA neurons, tonic DA activity facilitates LTD. Therefore, the predictive weight β_1 is modified to increase

and decrease as follows.

$$\Delta\beta_1(t) = \mu \cdot u_1(t) \cdot V'(t) \cdot burst(t) \cdot (limit - \beta_1(t)) - u_1(t) \cdot \epsilon \cdot tonic(t) \quad (3.18)$$

ϵ is the unlearning rate that determines the rate of weight decrease. Weight decrease occurs when there is input activity gated by the tonic DA signal.

Weight increase and decrease dependent on bursting and tonic DA activity is illustrated in Fig. 3.6. The burst spiking activity of the VTA is activated when its LH afferent becomes excited. Since tonic DA levels occur over a slower time course than DA bursts do (Grace, 1991; Floresco, 2007), tonic DA activity has been represented in Fig. 3.6 and in Eq. 3.10 as the low frequency component of the VTA DA neurons. While the burst is generated by LH activation on the VTA, tonic activity occurs due to the dis-inhibitory effect of the shell on the VTA via the mVP (Eq. 3.7). By dis-inhibiting the VTA, the shell influences tonic DA activity which activates DA D2-type receptors. It is proposed that tonic DA facilitates LTD (summarised in table 2.5(8)). This has been implemented in the model. In the following section, the shell circuitry will be modelled and its ability to influence tonic DA release will be realised.

Weight increase and decrease occurring in the adaptive system analogous to the NAc which comprises the shell and core subregions has been described so far. Evidence exist which suggest that the core plays a role in enabling reward predicting stimuli to mediate instrumental responding, while the shell mediates change in behaviour with respect to the incentive value of the conditioned stimuli (Floresco et al., 2008a). The adaptive system or NAc will be modelled according to the characteristics of the NAc subregions. Its shell unit updates the significance of the input stimuli it receives while its core unit learns to utilise reward predictive inputs to enable motor activity. The shell indirectly mediates the weight decrease by dis-inhibiting the DA neurons of the VTA.

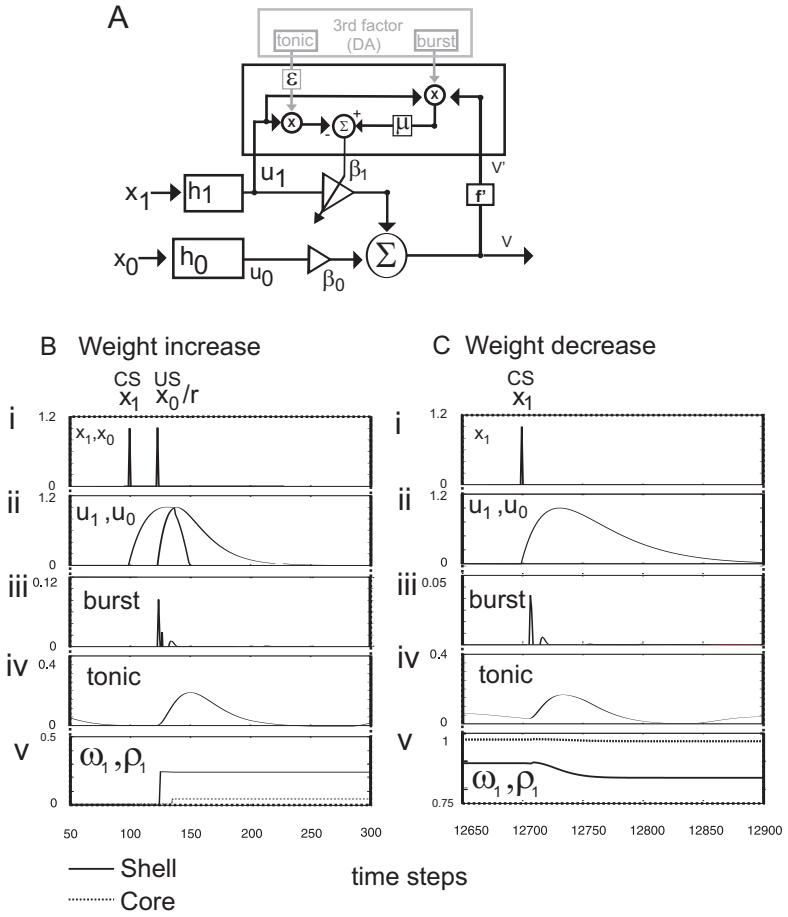

Figure 3.6: A) Circuit diagram illustrating how the plastic weight β_1 can increase and decrease. The weight increases according to the correlation between the pre-synaptic activity $u_1(t)$, post-synaptic activity $V'(t)$ and the third factor DA burst. The rate of weight increase is determined by the learning rate β. The weight decreases according to pre-synaptic $u_1(t)$ and tonic DA activity at a rate determined by ϵ. B and C) Signal traces illustrating how the weight increases and decreases respectively. i) The input x_0 and x_1 signals representing the US and CS. ii) The filtered x_0 and x_1 signals. iii) The burst. iv) The tonic activity B v) The weight increase dependent on the DA burst C v) The weigth decrease occuring due to tonic DA activity. Parameters: The h_{LP} frequency and q-factors are set to 0.01 and 0.51 respectively. The h_{HP} frequency and q-factors are set to 0.1 and 0.71 respectively. T=20 time steps. $\beta_0 = 1; \theta = 0.025$ (Thompson et al., 2009).

3.5.3 Modeling the NAc Shell Unit and Circuitry

So far it has been shown how the model agent processes inputs (x(t)) from the environment through input structures which feed into the adaptive system via plastic and fixed synapses (β). The plastic weights increase and decrease depending on the spiking activity of DA neurons. High frequency burst spikes are generated during LH activation while a much slower tonic DA transmission is influenced indirectly by the shell. The shell forms part of the adaptive system and as such comprises plastic weights.

The shell and its afferent and efferent structures are modelled and illustrated in Fig. 3.7, based on the shell connectivity described in the previous chapter. In this section, it will be shown how the shell dis-inhibits DA neurons and is capable of indirectly influencing cortical structures which innervate the core by inhibiting an area known as the medial ventral pallidum (mVP). The model of the shell circuitry aims to simulate this sub unit's ability to update values of stimulus that predicts rewards. Therefore, the shell requires information regarding reward availability. It obtains such information from the LH which becomes active when a primary reward is obtained (Kelley, 1999a). The shell is also innervated by limbic and cortical structures which provide information about the predictive inputs. In the model shell circuitry, the cortical units which when stimulated can maintain persistent activity are represented by the filtered predictive inputs OFC_j where $OFC_j = u_j$. These inputs summate with the LH onto the shell as follows:

$$shell = LH \cdot \omega_0 + \sum_{j=1}^{n} OFC_j(t) \cdot \omega_j \tag{3.19}$$

The shell associates the representations of predictive stimuli processed by the OFC to the rewards obtained which activate the LH. This association is represented in the weight ω, a specialised form of the weights β of the adaptive system to represent the shell weight. Activation of the predictive input that becomes associated with the LH produces activity in the shell which in turn inhibits the mVP, and reduces the inhibition on VTA neurons.

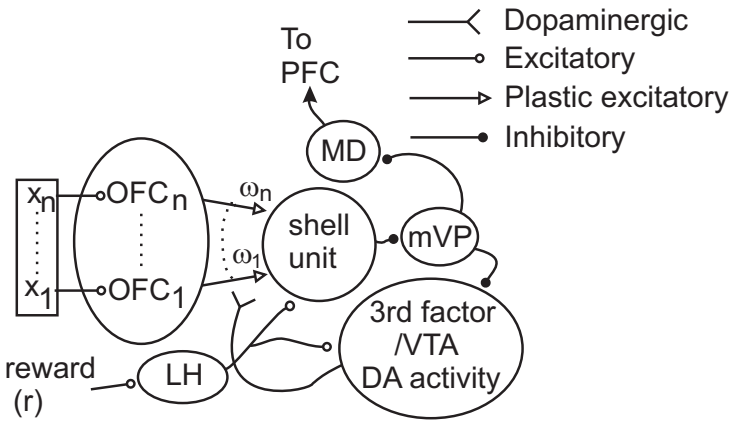

Figure 3.7: The shell circuitry which functions as an adaptive unit. The shell is the central hub that projects onto the mVP and along with the mVP also projects onto the VTA. It is innervated by the LH and **n** predictive inputs represented by the cortical structure indexed from OFC_1 to OFC_n. The LH projects onto the shell and VTA through a fixed pathway. The cortical units channel via their respective plastic weights ω onto the shell. These plastic weights can undergo both LTP and LTD. The shell influences the MD and VTA activities by dis-inhibition via the mVP.

Tonic DA produced in the absence of a reward mediates resultant LTD on synapse of the predictive input.

DA neurons transmit information in two activity states depending on whether or not a reward is obtained. The shell influences tonic DA activity by dis-inhibiting the dopaminergic neurons of the VTA. This is achieved by inhibiting the mVP which project sustained inhibitory activity on the VTA (Floresco, 2007) and MD as follows:

$$mVP(t) = \begin{cases} VP_{min} & \text{if } \frac{1}{1+\zeta \cdot shell(t)} < VP_{min}, \\ \frac{1}{1+\zeta \cdot shell(t)} & \text{otherwise.} \end{cases} \tag{3.20}$$

The VTA is dis-inhibited by the shell-mVP pathway as represented in Eq. 3.7. However the shell's ability to inhibit the mVP is capped at a minimum value indicated by VP_{min}. This might occur due to a limit on the population of

shell neurons that innervate the mVP. Therefore shells ability to influence the VP reaches a maximum. The shell also inhibits VTA neurons through a weak shell - VTA pathway weighted by η. Tonic DA generated enables the weights in the adaptive system to decrease. When rewards are absent, a lack of DA bursting activation in conjunction with increased tonic DA levels produced by the shell allows for resultant LTD to occur on plastic weights of the adaptive shell unit. Activity in the shell also generates a resultant initiation of a region known as the mediodorsal nucleus of the thalamus (MD) through this dis-inhibition. The MD forms part of the pathway that links the shell to the core, the second unit in the adaptive system.

$$MD(t) = \theta_{MD}(1 - mVP(t)) \tag{3.21}$$

Dis-inhibition of the MD provides positive feedback on cortical structures which project to the core. Through both the generation of tonic DA and the ventral pallido-thalamo-cortical pathway, the shell can indirectly influence the core activity. This pathway can be used to select updated input information which enable relevant behaviour initiated in the core unit of the adaptive system. The ability of the core in facilitating behaviour in response to cues that predict rewards is described in the following section.

3.5.4 Modeling the NAc Core Unit and Circuitry

The adaptive system is also specialised into the core as shown in Fig. 3.8. The core shares similar properties with the dorsal striatum and has been modified from the adaptive system to select actions based on the action selection model devised by Prescott et al. (2006). The core receives cortical information which provide preprocessed visual information representing the stimulus or cues available in the environment. The core unit enables the agent to learn to elicit behaviour in response to the cues processed by the PFC which suitably predict rewards. The predictive and US inputs are processed by the cortical innervation to the core are represented PFC_0 and PFC_j respectively. These

summate onto the adaptive core units through weighted channels as follows:

$$core(t) = PFC_0(t) + \sum_{j=1}^{n} PFC_j(t) \cdot \rho_j \tag{3.22}$$

ρ are the specialised core weights of the adaptive system's weight β.

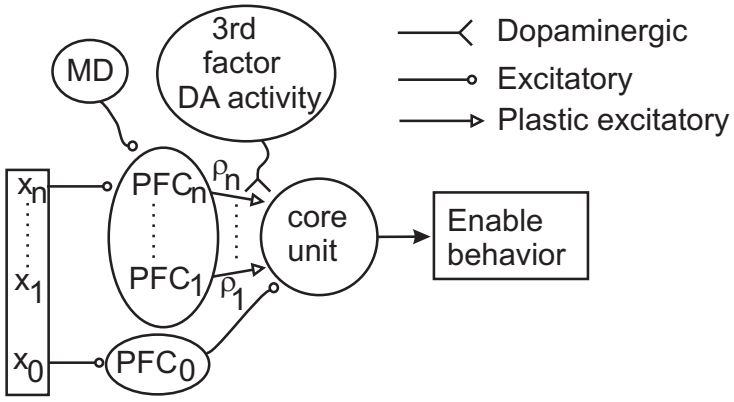

Figure 3.8: The core circuitry which functions as an adaptive unit that uses cues from the environment to enable behaviour. The core is innervated by the US input represented as PFC_0 and **n** predictive inputs represented by the cortical structure indexed from PFC_1 to PFC_n. The cortical units are also influenced by the shell activity via the MD which innervates the PFC_j units and the DA neurons of the VTA.

Plastic weights (ρ_j) change in the core according to the plastic weights of the adaptive system. As learning occurs, the core activity which facilitates behaviour becomes active in response to the cues in the environment that precede rewards. These cues are processed by the cortical inputs which project to the core. When the environment changes and these cues may no longer predict reward delivery, the agent must learn to adapt so that it can inhibit behaviour towards these currently irrelevant cues. The incentive values of stimuli representing the cues that once predicted rewards are updated in the shell of the adaptive unit. This updated information processed by the shell is capable of influencing and enabling adaptivity in the core by dis-inhibiting the MD. The MD innervates the cortical inputs that project to the core as

follows:

$$PFC_j(t) = u_j(t) + \theta_{MD} MD(t) \tag{3.23}$$

The core's ability to enable behaviour can be updated by MD activity through its cortical innervation. Therefore the shell, by dis-inhibiting the MD, is capable of influencing the core activity. This model is identified as an actor-critic model in chapter 5 whereby the shell represents the critic, while the core corresponds to the actor.

The following section describes how the input, reward and adaptive systems are incorporated to generate the full circuitry of the agent.

3.5.5 The Overall Model Circuitry

Fig. 3.9 illustrates the full circuit comprising the NAc shell and core and the reward and input systems. The NAc in the model is composed of one shell unit and **n** core units which acquire an association between predictive inputs and the reward and are capable of enabling **n** motor actions respectively. The model represented as an actor-critic model is discussed in chapter 5. The shell and core activities are represented by the information processed by the LH, OFC and PFC which summate onto their respective units as follows:

$$shell = LH\omega_0 + \sum_{j=1}^{n} OFC_j \cdot \omega_j \tag{3.24}$$

$$\begin{aligned} core\text{-}j &= [PFC_{0-j} \cdot \rho_{0-j} + \sum_{j=1}^{n} PFC_j \cdot \rho_j] \\ &- \sum_{k \neq j}^{n} \lambda \cdot core\text{-}k \end{aligned} \tag{3.25}$$

It is essential to point out here that **n** core units are represented in the model

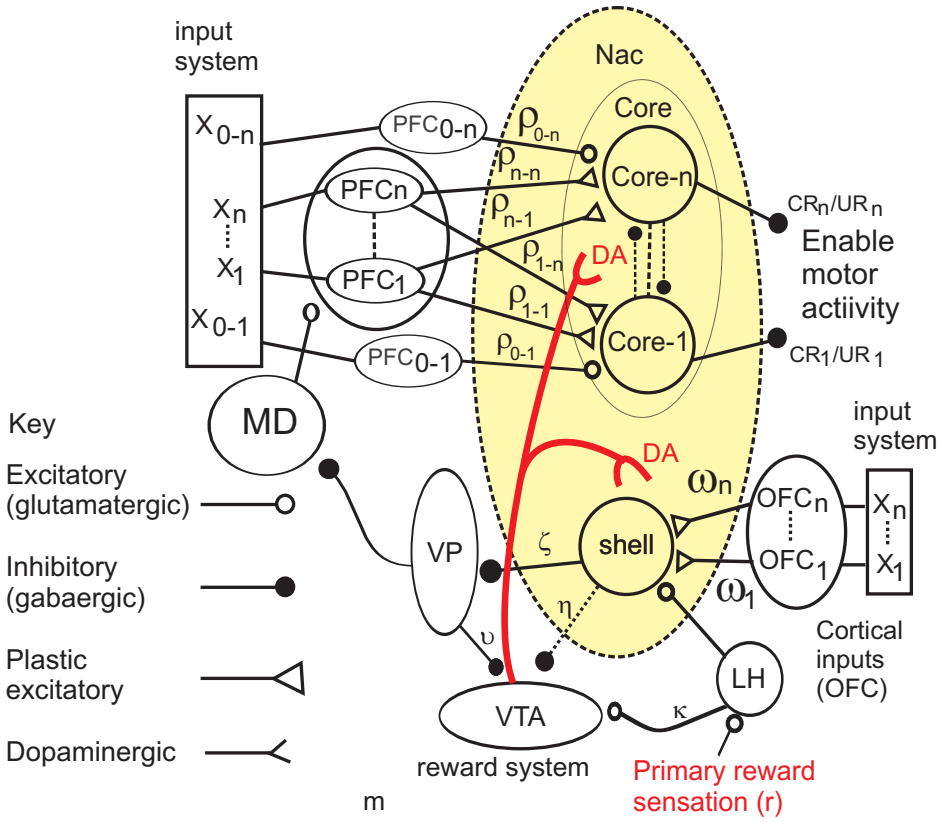

Figure 3.9: The NAc circuitry. The input system feed US and predictive inputs processed by the LH and OFC respectively to the shell and by the PFC to the core. The shell innervates the VTA and mVP. The mVP innervates the MD which projects onto the cortical neurons that feed into the core units. The core enables motor activity in response to activation from the predictive and US inputs. There are **n** core units indexed by j each innervated by **n** predictive inputs. For every predictive input is a US input indexed x_{0-j} which activates the $core_j$ and enables the respective UR indexed by j. An association can be learned between the predictive input (x_j) and US input (x_{0-j}) that occurs when a reward is obtained. This enables a respective CR indexed by j to be generated. The LH and shell influence VTA DA activity which feed onto the shell and core adaptive units and enable their corresponding plastic weights to change (Thompson et al., 2009). The $x_1...x_n$ inputs have been shown to feed into the core and shell. These inputs are general signals that originate from the environment. Their common source does not necessarily indicate that the shell and core units have the same inputs. However, any object from the environment might simultaneously stimulate different regions which link individually to the shell and core units.

each of which learn to enable **n** unique actions. For **n** inputs in the predictive pathway, there are **n** US inputs each of which feed into **n** individual core units and enable **n** distinct activities according to their output levels. It is assumed that each of these US inputs produces a common reaction. In addition, each core unit projects and inhibits to all other neighbouring core units by lateral inhibition via λ. This lateral inhibition ensures that the core unit with the strongest output suppresses all other respective core output and adapts a winner take all mechanism (Klopf et al., 1993; Suri and Schultz, 1999).

The number of predictive inputs are determined by **n** and for each predictive input, is a representative US input indexed between $0-1$ to $0-n$. All core units, although innervated by all the predictive inputs can be identified by their distinct US input. The predictive input which correlates with the US input is capable of activating its respective core unit. The predictive inputs that innervate the core are influenced by the shell through the shell-VP-MD pathway.

The shell is innervated by OFC neurons which have also been identified as predictive inputs. These OFC inputs are capable of enabling the shell if their activity correlates with the LH which becomes active when a reward is obtained from the environment. The LH innervates the VTA whose DA neurons undergo two kinds of activity states. The DA neurons gate plasticity in the shell and core units.

Plasticity is modelled as follows according to Eq. 3.18 :

$$\begin{aligned}\beta_j(t) \leftarrow \beta_j(t) + [\mu \cdot u_j(t) \cdot V'(t) \cdot burst(t) \cdot (limit - \beta_j(t))] \\ - \epsilon u_j(t) \cdot tonic(t)\end{aligned} \tag{3.26}$$

Such that the weight increases to a maximum value bounded by "*limit*". Weight changes in the shell accordingly as:

$$\begin{aligned}\omega_j(t) \leftarrow \omega_j(t) + [\mu_{shell} \cdot OFC_j(t) \cdot shell'(t) \cdot burst(t) \cdot (limit - \omega_j(t))] \\ - \epsilon_{shell} \cdot OFC_j(t) \cdot tonic(t)\end{aligned} \tag{3.27}$$

Similarly, the weight change in the core is as follows:

$$\rho_j(t) \leftarrow \rho_j(t) + [\mu_{core} \cdot PFC_j(t) \cdot core'(t) \cdot burst(t) \cdot (limit - \rho_j(t))] \\ - \epsilon_{core} PFC_j(t) \cdot tonic(t) \tag{3.28}$$

In the model, the core is responsible for acquiring an association between rewards and the cues that predict them and enabling behaviour in response to these cues. The shell also acquires an association between cues that predict rewards and the rewards and in addition mediates an adjustment in behaviour when the incentive value of the stimulus predicting the reward changes and no longer predicts the reward.

When a reward is omitted and a CS which predicted the reward no longer precedes reward delivery, the shell mediates change in response towards that stimuli by influencing the tonic DA release. Tonic DA is released proportional to the shell activity. Initially the shell activity will be higher and slowly decreases as tonic DA activity enables weight decrease in the shell. This will eventually result in reduced tonic DA activity. results in a decrease in tonic DA release. In addition, the shell influences the core's ability to enable behaviour via the shell-mVP-MD-PFC-core pathway. It will be shown how the shell enables change in behaviour when contingencies change by quickly learning which stimuli predict and no longer predict rewards. On the other hand the core which learns to enable behaviour in response to the CS that precede rewards is modelled not to quickly unlearn such associations. This is because such learning and then unlearning of enabling behaviour have been argued by both (Rescorla, 2001) and (Bouton, 2002) not to occur during extinction. In order to support this theory and ensure that learned associations which enable behaviour are not eliminated one major assumption is made about DA transmission in the shell and core. The previous chapter summarised studies which observed distinct DA transmission in the shell and core during appetitive behaviour. In the present model, DA transmission mediates LTP and LTD differentially. The rate at which the plastic weights in the shell (ω_j) and in the core (ρ_j) change are assumed to occur at different

rates whereby the rate at which the weights decrease in the shell occurs at a greater rate than in the core. This ensures that information about cues that predict rewards are quickly updated in the shell but not quickly unlearned by the core.

In the following sections, the mechanism by which the model functions is observed using a set of open-loop simulation experiments.

3.6 Simulations of Classical Conditioning

Classical conditioning simulations are conducted in which delta pulses which represent predictive and US inputs are fed into the input system of the NAc or adaptive unit. These simulation experiments have been carried out to demonstrate how DA neurons predict reward availability and omission. During these experiments the shell's ability to update information about reward predictive cues will also be illustrated. In addition, the model's ability to account for some classical conditioning effects will be observed.

In these simulations, a set of delta pulses are presented which may or may not correlate with the delta pulse that represents the reward. The generalised NAc circuitry illustrated in Fig. 3.9 is used. The experiments show how the circuitry processes and adapts to inputs which correlate with the delivery of rewards. The first simulation experiment will be used to illustrate how the reward system's DA transmission encodes the presence or absence of unexpected rewards and adapts as reward delivery becomes available. The second set of simulations will show how the model accounts for certain basic features of Pavlovian conditioning. Such basic features include acquisition, extinction, blocking, overshadowing and the interstimulus interval (ISI) effects.

Learning and unlearning occur in the adaptive system depending on whether or not a reward is available. The next section illustrates how the VTA reward system encodes reward availability and predictability. The shell influences the reward system and updates information when a CS predicts a reward,

and when the reward is omitted.

3.6.1 Simulating Tonic-Phasic Dopaminergic Activity

This section demonstrates how the VTA tonic and phasic DA activity encodes reward availability. Simulations were coded in C++ using a Linux workstation. Filters were implemented as time-discrete IIR filters. The simulation parameters implemented in the experiments are provided in table in appendix B.

The mechanism for detecting reward availability is utilised by the agent to adapt and therefore respond accordingly. In addition to illustrating how the model simulates DA activity, it will be shown how the agent obtains a CS-US association when they are presented together and how the acquired association changes during extinction.

DA neurons initially fire short phasic bursts of activity during the presentation of unexpected primary rewards. When expected rewards are omitted an absence of phasic DA activity is observed. As learning becomes established these DA neurons fire less significantly during the receipt of reward and fire during the presentation of an originally neutral stimulus that consistently precedes the reward (Schultz et al., 1993; Schultz, 1998). DA neuron activity is obtained directly from the influence of its afferent innervations.

Fig. 3.10 illustrates how the DA neurons in reward system of the model performs in accordance with these findings. Fig. 3.10 shows the CS or x_1 and US or x_0 signals, the LH, VTA, burst, tonic and shell (ω_1) and core (ρ_1) weights. Fig. 3.11 shows the CS or x_1 signals, the LH, VTA, burst, tonic and shell (ω_1) and core (ρ_1) weights when the reward or US is omitted. During acquisition, between the time steps 0 to 12300, delta pulses x_1 (CS), x_0 (US) or r (reward) are consistently fed every 300 time steps into the input system of the circuit (Fig. 3.9) and represent the CS, US and reward. Only one CS to US pair are used to represent acquisition. These are processed by the cortical inputs of the PFC and OFC and the LH represented by u_1 and u_0

or LH. Fig. 3.10 shows three trials at different times during acquisition. The CS and US are paired, between the time steps 50 to 300 (at the first trial), 2750 to 3000 (at the tenth trial) and 12050 to 12300 (at the 41st trial). The x_0 or r is presented at an interval of 20 time step after the CS. Between the time steps of 50 to 300, the CS or x_1 signal is initially presented followed by the US or x_0. These signals are processed to represent the cortical and LH input. The OFC inputs potentially have access to the VTA via the shell-mVP pathway. This means that the CS inputs have the potential to activate the VTA and thus generate a DA burst of activity at the CS onset. However, the ω_1 weight between the CS inputs channelled by the OFC and the shell are initially set at zero. The LH input activates the VTA whose highpass and lowpass filtered components generate the resultant burst and tonic activities. The burst generated enables the plastic weights in the shell (ω_1) and core (ρ_1) to increase. During the time steps of 2750 and 3000 the shell weight increases and the predictive CS input generates resultant activity in the shell which dis-inhibits the VTA. This enables the VTA activity to occur during CS presentation. A burst is generated in the event of the CS. The shell weight then increases in later trials such that its direct inhibition on the VTA becomes stronger than the dis-inhibition via the mVP pathway. When this happens the DA burst occurring during the LH activation gets suppressed and decays due to the shell's greater direct inhibition on the VTA than the VTA's activation via the LH and Shell-VP-VTA pathways. The DA burst during the CS onset increases because the rate of change of activity at the CS onset is greater than the rate of change of activity at the US onset. The filtered inputs u represent the cortical projections to the NAc. When the primary reward is obtained, the cortical activity is slowly silenced. As the reward is delivered, the filtered predictive u_1 input which contributes to the NAc activity decays. This decay reduces the population of tonically activated DA neurons and therefore the resultant weight decrease enabled via this process is minimal.

In the current model, the input system to the NAc include the cortical innervations which until a reward signal is produced, generate prolonged activity.

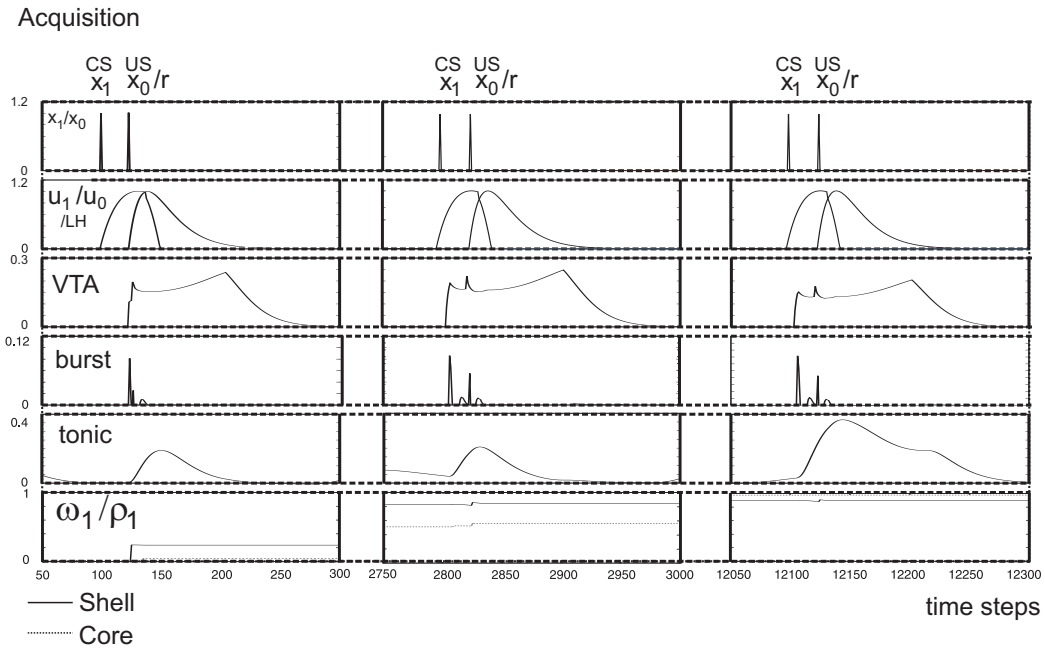

Figure 3.10: Simulating DA neurons and the two characteristic tonic and burst activities. Acquisition is observed by the increase in $\omega_1 or \rho_1$ Parameters: The h_{LP} frequency and q-factors are set to 0.01 and 0.51 respectively. The h_{HP} frequency and q-factors are set to 0.1 and 0.71 respectively. T=20 time steps. $\beta_0 = 1; \theta = 0.025$ $\mu_{shell,core} = 0.5$; $\epsilon_{shell} = 0.01$; $\epsilon_{core} = 0.0005$.

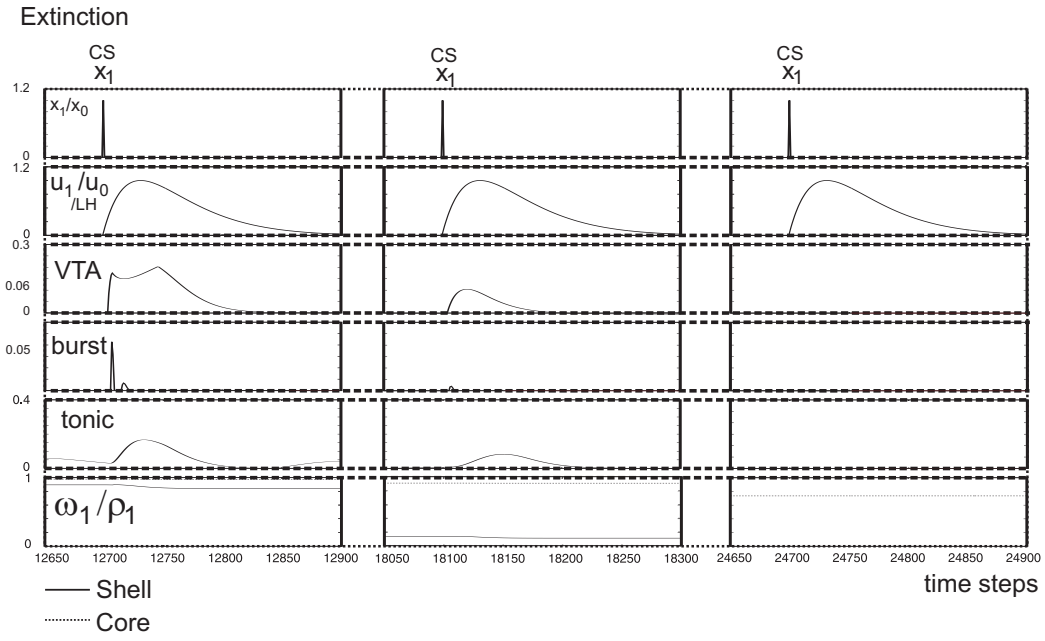

Figure 3.11: Simulating DA neurons and the two characteristic tonic and burst activities. Extinction is observed by the decrease in $\omega_1 or \rho_1$ Parameters: The h_{LP} frequency and q-factors are set to 0.01 and 0.51 respectively. The h_{HP} frequency and q-factors are set to 0.1 and 0.71 respectively. T=20 time steps. $\beta_0 = 1; \theta = 0.025$ $\mu_{shell,core} = 0.5$; $\epsilon_{shell} = 0.01$; $\epsilon_{core} = 0.0005$.

Therefore, the input system that innervate the shell, are modelled as the OFC which maintain persistent activity for a delayed period or until a reward is obtained. During extinction after the time step 12650 Fig. 3.11, the reward is no longer presented. However, as the CS is presented, the cortico-shell activation results in extended tonic DA transmission and a longer duration for LTD. The DA burst generated only during the CS onset is insufficient to result in an overall weight increase. As the weight decreases, the shell's influence on the VTA via the mVP is reduced. The weight decrease results in a reduction in the shell activity which also drives the VTA. The DA burst or tonic activity eventually disappear completely (Fig. 3.11 time steps 14100 to 14400). The DA bursts generated during the CS events are useful for enabling higher order conditioning while the tonic DA levels are useful for enabling the weights to decrease.

In this section, it has been shown how CS-US pairings have generated DA bursts which enable resultant weight increase and the acquisition of CS-US associations. The next section addresses how the model performs during a variety of classical conditioning processes. This commences with acquisition and extinction.

3.6.2 Acquisition and Extinction

Acquisition and extinction can be represented as follows

$$CS+ \rightarrow UR \Rightarrow CS0 \rightarrow CR \tag{3.29}$$

$$CS0 \rightarrow noCR \tag{3.30}$$

The single arrows indicate pairings, while the double arrows illustrate what happens after a duration of repeated pairings. Acquisition occurs when a CS is paired with a US ($CS+$) and an association is established between the CS and US such that a CR occurs when the CS is presented. Extinction occurs when the CS that elicits a CR is not presented with a US ($CS0$) and the CR is no longer generated. The weight development in the shell

and the core during acquisition and extinction is shown in the lowest panel of Fig. 3.10 and Fig. 3.11 respectively. The shell and core weights and CR represented by core output during acquisition and extinction are shown in Fig 3.12. The unshaded region in Fig 3.12A, B and C illustrates how the shell weight (ω), core weight (ρ) and magnitude of the core output (CR) increase during acquisition when the CS and US are paired. During omission when the CS is not paired with the US, the weights and CR decrease. During acquisition, the weight in the core develops such that the core enables activity (CR) in response to the CS. In addition, the shell facilitates this core activity by dis-inhibiting the MD which projects onto the cortical afferents of the core. In this way, the shell activity contributes to the CR magnitude. It can be seen in Fig 3.12 A and B how the shell and core weights and the CR increase and decrease in synchrony during acquisition and extinction. However, during extinction, the core weights do not decrease to 0. This ensures that the association between the CS and US is not completely destroyed. The shape of the CR magnitude during acquisition Fig 3.12 C is similar to the desirable S-shape curve observed in empirical data (Patterson et al., 1977; Gibbs et al., 1978). In addition, the shape of the curve during acquisition and extinction resemble the pattern observed in empirical data showing acquisition and extinction (Schneiderman and Gormezano, 1964; Smith et al., 1969; Frey and Ross, 1968). Although the CR do not decrease to a value of 0 during extinction, a threshold value could be set over which the CR must reach before a response becomes apparent. In this case, the threshold value could be set to a value of 0.5 as indicated by the vertical line in Fig 3.12 C. Below this value, the CR is not apparent in behaviour.

3.6.3 Interstimulus-Interval Effects

The influence of ISI has been identified to be an essential indicator of performance in classical conditioning (Schneiderman and Gormezano, 1964). It indicates how a CR is dependent on the temporal interval between the CS and US pairings. An ideal response level demonstrates the following properties

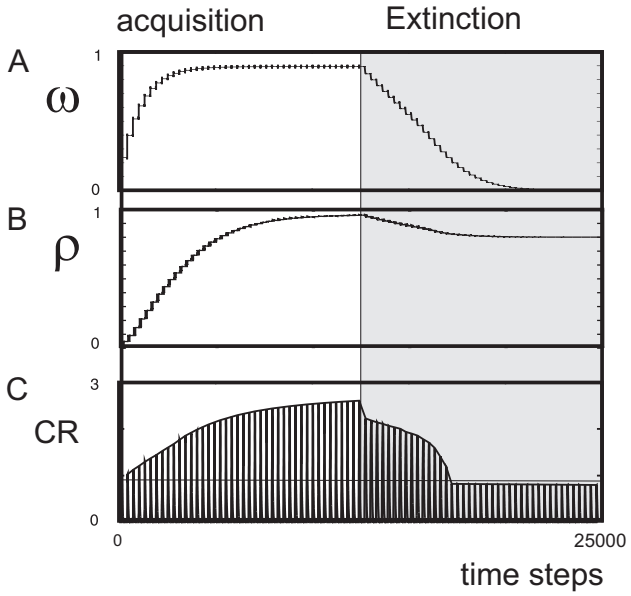

Figure 3.12: Simulation run over a duration of 40000 time steps illustrating acquisition and extinction in A) ω, the shell weight, B) ρ, the core weight and C) the conditioned response CR magnitude which represents the core output. During the first half of the simulation run, a CS1 is presented followed by a US presentation at a time step T=20 after the CS1. The weight development increases as the CS and US are consistently paired. During extinction, the second half of the simulation run, the US is omitted generating a resultant decrease in the ω_1 weight. Parameters are presented in appendix C in table C.1.

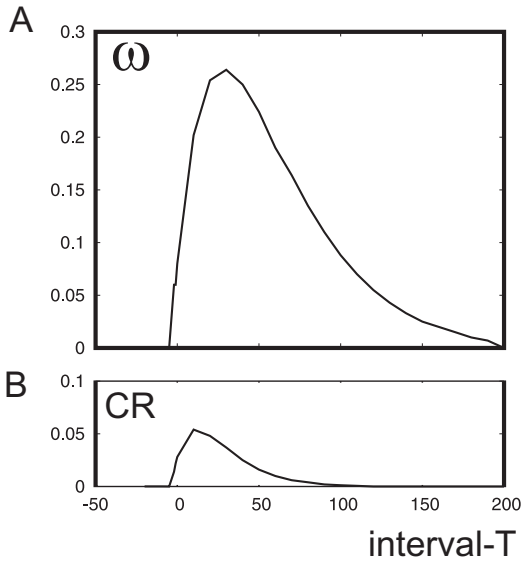

Figure 3.13: A) The weight change curve of the shell and B) the CR magnitude as a function of T, the interval between the CS and US presentation illustrate the model's ISI dependency. Parameters are presented in appendix C in table C.1.

(Balkenius and Morén, 1998):

- Zero response level at negative ISI intervals i.e. when the US precedes the CS.
- A single maximum response level peak is observed at small positive ISIs.
- As the ISI increases, an asymptotic decline in the response level results.

The model's ability to reproduce an ISI-curve is illustrated in Fig. 3.13. The figure shows both the change of the shell weight ω (Fig. 3.13A) and the maximum CR (Fig. 3.13B) as a function of the delay T between the CS and the US. Here it can be seen that the shapes of both the weight change curve and CR resemble the ISI curve of the adapted empirical data presented by Balkenius and Morén (1998) of the nictitating membrane experiments

conducted by Smith et al. (1969); Schneiderman and Gormezano (1964). The model's ability to account for overshadowing is discussed next.

3.6.4 The Overshadowing Effect

Overshadowing which was first reported by Pavlov (1927) but which has been demonstrated in later studies (Kamin, 1969), occurs when conditioning on compound stimuli results in each stimulus obtaining weaker associative strengths than if they were individually paired with the same US (Pearce, 2008). Overshadowing can be seen when two stimuli are presented together with a US. The strength of CR produced is relatively weaker than the strength of CR that would have been generated if they had been conditioned individually with the US Pearce (2008).

Overshadowing is tested in the model by running two simulations (A and B) in parallel for comparison. In the first simulation (A), A pair of delta pulses are presented 20 time steps (T=20) prior to the US $(CS1_A\ CS2_A)+$. This occurs every 300th time step for a duration of 10000 until an association is formed. The response strength is observed by presenting the $CS2_A0$ and measuring the $CR2_A$ magnitude. Similarly in the second set of simulations (B), only one delta pulse is presented with the US ($CS2_B+$) every 300th time step for a duration of 10000. Afterwards the response strength is observed by presenting the $CS2_B0$ and measuring the $CR2_B$ magnitude. The test is demonstrated thus:

$$(CS1_A CS2_A)+ \Rightarrow CS2_A \rightarrow CR2_A \tag{3.31}$$

$$CS2_B+ \Rightarrow CS2_B \rightarrow CR2_B \tag{3.32}$$

The maximum CR2 in both simulations A and B are obtained by observing the magnitude of the core-1 and core-2 units output activities. These are represented in Fig 3.14. It can be seen that the maximum $CR2_B$ value is greater than the maximum $CR2_A$ value indicating that overshadowing occurred during compound conditioning in simulation A.

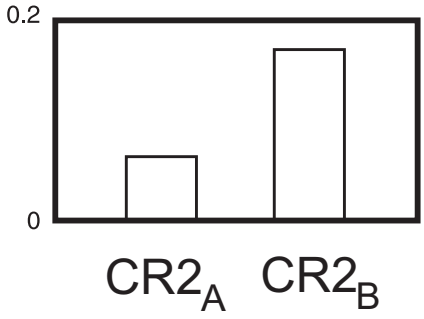

Figure 3.14: The overshadowing effect $CR2_A$ generated due to compound conditioning is smaller than $CR2_B$ generated from the single CS2-US conditioning. Parameters are presented in appendix C in table C.1.

The overshadowing effect shows that that the associative strength of stimuli conditioned individually with the US is different to the associative strength obtained when the stimuli are combined. Acquisition of each core unit is achieved in the model through 3-factor learning whereby the 3 factors include pre-synaptic activity (PFC input), post-synaptic activity (core output) and DA burst (Eq. 3.28). During overshadowing, the output of each core unit is inhibited by the neighbouring core connectivity (Eq. 3.25) such that the post-synaptic activity is inhibited and the resultant weight change is reduced. The model has successfully shown that overshadowing as observed in animal behaviour can be produced. Another paradigm which the model is capable of simulating is the blocking effect.

3.6.5 The Blocking Effect

Blocking (Kamin, 1969) is a classical conditioning phenomenon which leads to the suggestion that the unpredictability of the US influences conditioning (Pearce, 2008). During blocking, a CS1 is paired with a US ($CS1+$) until an association is learned. Afterwards, the CS1 is presented in combination with a stimulus CS2 and paired with the US (($CS1CS2)+$). The result is that the original training with CS1 alone blocks the learning of the association between CS2 and the US. Blocking is summarised as follows:

$$CS1+ \rightarrow UR \Rightarrow CS10 \rightarrow CR1 \tag{3.33}$$

$$(CS1CS2)+ \Rightarrow CS20 \rightarrow noCR2 \tag{3.34}$$

The blocking effect is tested in the model by running two simulations (A and B) in parallel for comparison. In the first simulation (A), a delta pulse is delivered 20 time steps (T=20) prior to the US delivery ($CS1_A+$). This occurs every 300th time step for a duration of 10000 time steps until an association is formed. After this a second delta pulse is delivered in conjunction with $CS1_A$ at 20 time steps prior to the presentation of the US (($CS1_ACS2_A$)+). Again, this occurs every 300th time step for a further 10000 time steps. In the second set of simulations (B), a delta pulse is delivered with the US omitted ($CS1_B0$) every 300 time steps for a duration of 10000. Afterwards a second delta pulse is delivered in conjunction with $CS1_B$ each at 20 time steps prior to the presentation of a US every 300 time steps for a further duration of 10000. The blocking effect is summarised by the equations:

$$CS1_A+ \rightarrow (CS1_ACS2_A)+ \Rightarrow CS2_A \rightarrow CR2_A \tag{3.35}$$

$$CS1_B0 \rightarrow (CS1_BCS2_B)+ \Rightarrow CS2_B \rightarrow CR2_B \tag{3.36}$$

The CR2s represent the associative strengths developed between the $CS2_A$ and the US and $CS2_B$ and the US respectively in the two simulation groups A and B. These responses are represented in Fig. 3.15 which shows that the $CR2_B > CR2_A$. The acquisition of the $CS2_A$-US association is blocked by the initial acquisition of the $CS1_A$-US association. In addition the compound conditioning occurs during simulation B and once again the overshadowing effect as described in the previous section is observed in the $CR1_B$ and $CR2_B$ magnitudes.

Just as in overshadowing, blocking is achieved due to the output of each core

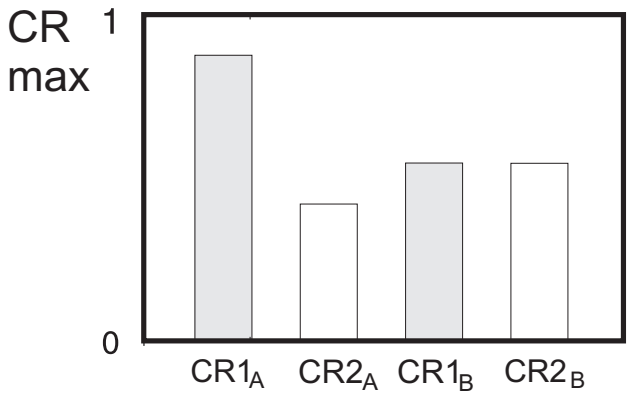

Figure 3.15: The blocking effect Parameters are presented in appendix C in table C.1.

unit inhibiting neighbouring core elements (Eq. 3.25) and therefore the magnitude by which the weights increase. This results in a reduced magnitude of the core activities. The next section illustrates how the model performs during rapid reacquisition.

Acquisition of each core unit is achieved in the model through 3-factor learning whereby the 3 factors include pre-synaptic activity (PFC input), post-synaptic activity (core output) and DA burst (Eq. 3.28). During overshadowing, the output of each core unit is inhibited by the neighbouring core connectivity (Eq. 3.25) such that the post-synaptic activity is inhibited and the resultant weight change is reduced

3.6.6 The Reacquisition Effect

The reacquisition effect is apparent when a previously extinguished CR recurs much quickly than its initial appearance (Pavlov, 1927) It has been tested by subjecting the model to acquisition and extinction four times and comparing the response rate of the model during the acquisition and reacquisition stages. Rapid reacquisition Fig. 3.16, is illustrated by observing the CR of the model over the duration of the test. The initial CR magnitude generated during acquisition are indicated by the arrows in Fig. 3.16. The maximum value

of the CR at the beginning of the reacquisition is larger than the maximum value of the first CR during acquisition. Therefore, the model shows rapid reacquisition. This is achieved because unlike the shell weights, the core weights do not quickly reduce to 0. And a CR can instantly be reproduced as described in section acquisition.

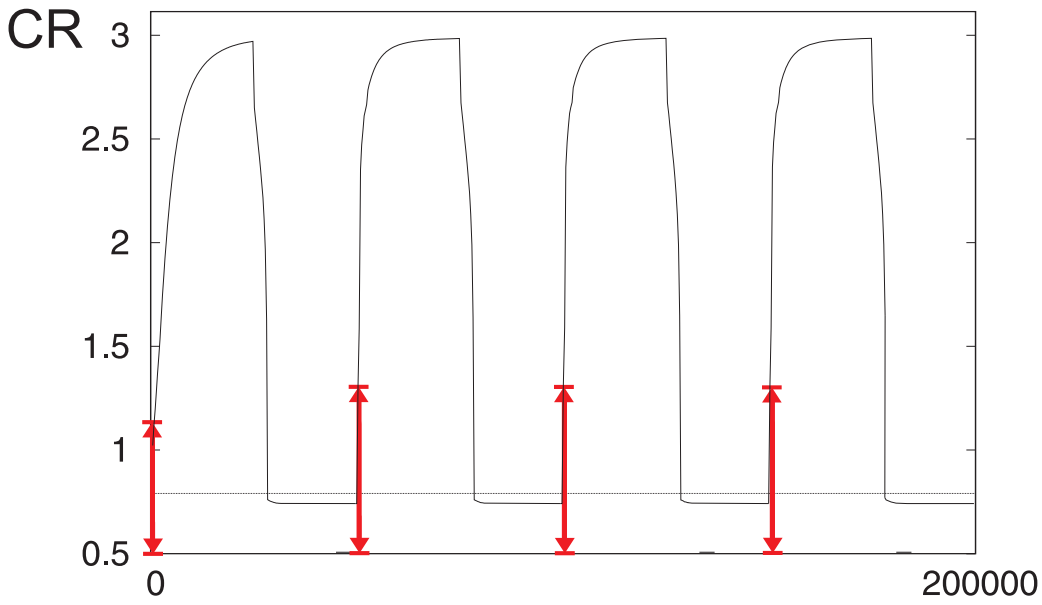

Figure 3.16: Initial acquisition and 3 consecutive reacquisition. Parameters are presented in appendix C in table C.1.

3.7 Concluding Remarks

In this chapter, the environment and agent were introduced and tested as an open-loop system using simulations that represented classical conditioning paradigms and whereby the response of the system does not determine whether a reward is delivered. A model of the circuitry surrounding the NAc has been integrated into the agent which obtains and processes information from the environment. The processes by which the agent is able to use cues from its environment to obtain rewards have been formalised and observed in open-loop simulations. In these simulations, the agents ability to encode

reward availability have been observed. A variety of classical conditioning paradigms have been tested by the model.

In the following chapters, the model is tested in scenario behavioural experiments. The agent will be required to perform reward seeking tasks by using landmarks available in the environment as cues. The agent's ability to learn, adapt and perform behavioural flexibility will be tested and demonstrated in simple reward based learning, reversal and secondary conditioning scenario simulation experiments.

Chapter 4

The Closed-Loop Behavioral Experiments

4.1 Introduction

In chapter 3, a computational model of the nucleus accumbens (NAc) circuitry was developed. Its adaptability in open-loop conditions was also observed. In this chapter, the model will be tested in closed-loop behavioural scenario reward seeking experiments. In the closed-loop experiments, the reward delivery is contingent on the response generated by the agent. These behavioural experiments are used to obtain a basis for which the model can be analysed against the NAc circuitry's role in acquisition and reversal of reward based behaviours. The model's capability to learn in an environment will be formalized in a simple reward seeking task. Following the reward seeking experiment, the shell and core of the model will be lesioned and tested in the behavioural reward seeking tasks. The results generated will be compared against empirical results obtained from the shell and core lesion studies conducted by Parkinson et al. (2000).

In addition to the basic reward seeking task in which the agent must learn to approach a landmark containing a reward from a distance, the model will also

be tested in a reversal learning food seeking task. During reversal learning, the agent learns to discriminate between the CS+ and CS- which are stimuli that do and do not predict rewards respectively. After the agent has acquired an association, the contingency is reversed such that the CS+ which initially predicted the reward becomes the CS- and no longer precedes the reward and vice versa for the original CS-. The agent inhibits behaviour towards the now irrelevant conditioned stimulus (CS) and learns the new association.

These experiments combined show how the model is capable of learning and adapting in rather simple and changing environments. Each behavioural setup will initially be summarised followed by a brief description of how the model processes information to complete the task. The corresponding simulation run and the results obtained for each scenario task will be shown. In the next section, the implementation of the circuitry into an agent, environmental setup and agent-environment interaction is described.

4.2 The Behavioral Experiments

Three sets of behavioural reward seeking experiments are carried out in this chapter which are used to appreciate the NAc's functionality in mediating reward seeking behaviours. The first set of experiments are used to demonstrate how the signals are processed by the model NAc circuitry during a reward seeking behaviour. The second set of experiments compares the performance of the model when subject to shell and core lesions to empirical data which show the effects of shell and core lesioned agents during conditioning. The final set of experiments show how the model is capable of performing reversal learning by inhibiting behaviour which although was once useful no longer predicts a reward.

4.2.1 The Environment and the Agent

The computational model is integrated into an agent and tested in environments which are simulated on a Linux platform using an opensource open dynamics engine (ODE)[1] programmed in C++. The simulated experimental setup consists of the environment in which is contained an agent, a number of landmarks and a reward embedded within a landmark. The agent is capable of approaching the landmarks which elicit individual proximal and distal signals.

The simulated environment for testing learning is shown in Fig. 4.1A. This octagonal environment comprises the agent, a yellow and a green landmark and a reward embedded inside one of the landmarks. In the reward seeking behavioural experiment Fig. 4.1, the agent explores the environment and must learn to find the 'reward' (Porr and Wörgötter, 2003; Verschure et al., 2003; Thompson et al., 2008) located within a landmark. Only one landmark at a time can contain the reward. The agent's starting point at the beginning of the experiment or once it has come in contact with the reward is at a random point on the line located in the centre of the environment equidistant to either of the landmarks.

The agent is shown in Fig. 4.1B. It contains color sensors which detect the colored landmarks and rewards and touch sensors for detecting the walls Fig. 4.1B. The agent detects the landmark through distal and proximal sensors. Fig. 4.1C shows how a landmark X as an example, elicits signals which the agent can detect as proximal signals when the agent is located in close vicinity to the landmark and as distal signals elicited when the agent is at a distance from the landmark.

The proximal signals are filtered (u-proximal) and weighted ($\rho_{X-proximal}$) with fixed values greater than 0. They function as soft decision makers which gate the agent's action-subsystem. The action subsystem is modelled as a Braitenberg vehicle (Braitenberg, 1984) which comprises sensors that activate

[1] http://www.ode.org/

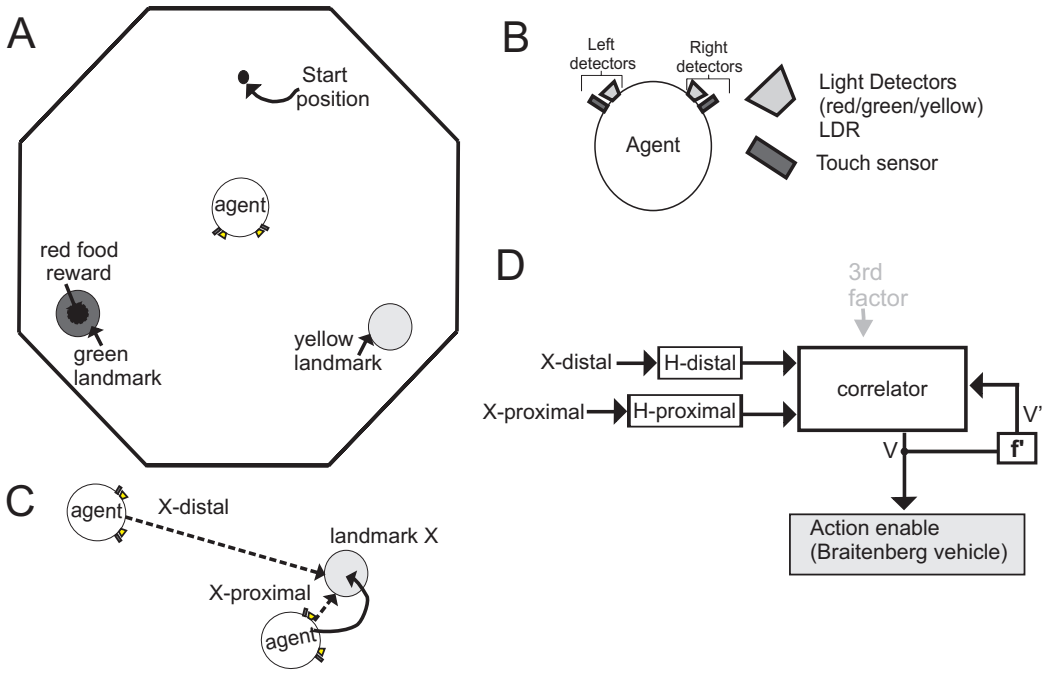

Figure 4.1: A) The environment containing the agent, a yellow and green landmarks and a reward embedded in the green landmark. B) The Agent with left and right light detectors and touch sensors. C) The proximal (X-proximal) and distal (X-distal) signals of the landmark X detected by the agent. X represents either the yellow (Y/y) or green (G/g) landmark in the environment. D) The X-proximal and X-distal signals through their ρ_0 and ρ_1 weights respectively, are capable of gating the action subsystem which is implemented as a Braitenberg vehicle. (Thompson et al., 2009)

effectors via a gated channel that is influenced by the proximal signals. This means that when the signal is active it is capable of immediately enabling the Braitenberg vehicle. The proximal sensors generate a set of signals related to the unconditioned stimulus (US) which trigger the pre-wired reflexes and directly enable a reaction (UR) when the agent is in close proximity with the landmark Fig. 4.1D. This attraction behaviour has been interpreted in the previous chapter as exploratory behaviour. The naïve agent can only navigate to the landmark when it is at this proximal distance to the landmark. The agent can also detect the landmarks from a distance through the distal sensors which generate the CS representations. The distal signals ($\rho_{X-distal}$) are also

filtered (u-distal) and weighted with with plastic weights. These signals also have the ability to facilitate the action subsystem in a similar manner to the proximal signals however, if and only if their plastic weights are not equal to zero. In the naïve agent, these weights are originally set to zero indicating that the agent does not have pre-wired conditioned reflexes. Initially, before learning the agent does not approach the landmark from the distance instead, it learns to approach the landmarks from the distance depending on whether or not the landmark contains a reward.

Learning occurs when the reflex reaction (UR) which results in the delivery of a reward, correlates with the distal signal representation (CS) so that the agent is capable of targeting the reward when the distal signals are elicited. The plastic weights change depending on the correlator in Fig. 4.1D which correlates the distal with the proximal signals as the agent explores the environment and finds the reward. Therefore, the agent learns an association between the reward and the landmark that contains it. The distal signals from other landmarks can also be fed into the network in Fig. 4.1D and utilized in an identical manner to the X-distal signal. This means that the signals from the surrounding landmarks integrated into the network can also drive motor activity just as the distal signals from the landmark X can. The reflex x_0, predictive x_1 and r signals utilised in the open-loop experiments in the previous chapter are reinterpreted as the proximal, distal and reward signals respectively.

Upon learning, the plastic weights of the distal signals change and enable the agent to approach the landmark containing the reward from a distance. Information is processed in the circuitry as described in the previous chapter. The information flow in the model which is processed according to the signals obtained from the specialized environment is briefly summarized next.

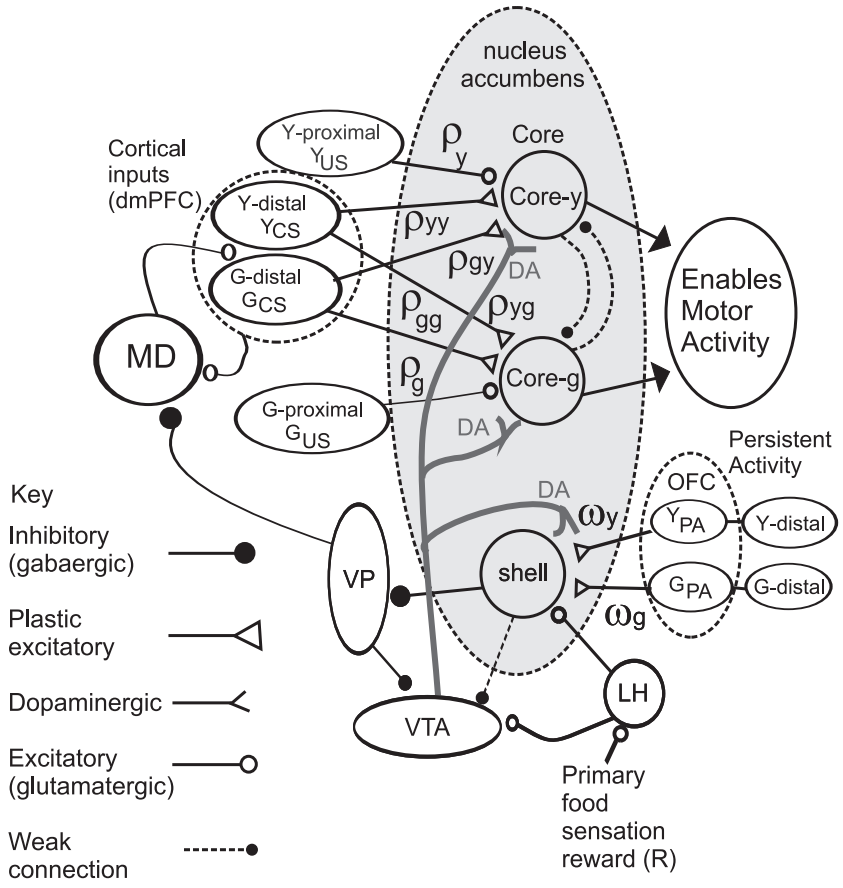

Figure 4.2: The full limbic circuitry model adapted for the behavioural reward seeking task. Distal and proximal signals from the yellow (Y) and green (G) landmarks represent sensor inputs feeding into the respective dorsomedial prefrontal cortex (Y_{CS} and G_{CS}) and the orbitofrontal cortex (Y_{PA} and G_{PA}). The cortical inputs innervate the NAc core and shell units. Primary food rewards activate the lateral hypothalamus (LH) which projects to both the ventral tegmental area (VTA) and the shell. The shell innervates the medial ventral pallidum (mVP) and the ventral tegmental area. The ventral pallidum innervates the mediodorsal nucleus of the thalamus (MD). The core units use cortical activities to mediate motor behaviours. These cortical afferents to the core are indirectly influenced by the shell via the mVP-MD-PFC pathway. The shell also influences the VTA which releases DA and mediates plasticity in both the core and the shell units. (Abbreviations: LH, lateral hypothalamus; PFC, prefrontal cortex; OFC, orbitofrontal cortex; VTA, ventral tegmental area; mVP, ventral pallidum; MD, mediodorsal nucleus of the thalamus; PA, persistent activity) (Thompson et al., 2009)

4.2.2 The Agent Model

The full model circuitry adapted to process information from the environment is illustrated in Fig. 4.2. The circuitry is composed of the biologically relevant input, processing and motor regulatory structures capable of influencing behavioural food seeking tasks. The signal processing pathway in the model commences from the cortical input of the PFC to the NAc to activate the motor system or the VTA neurons. The simulated circuitry comprises the NAc's distinct shell and core subunits as the central hub. The OFC region of the PFC innervates the shell and processes information representing the visual inputs from the landmarks. On the other hand, the dmPFC innervates the core and provides preprocessed visual information representing the landmarks or reward.

The output of the core gates each action subsystem as implemented by the basal ganglia in Prescott et al. (2006) and comprises of sub nuclei each of which enables one particular motor activity or behavioural response. In this case each behavioural response corresponds to the attraction behaviour controlled by the Braitenberg vehicle. The two different landmarks that can be approached are activated by two individual core-y and core-g nuclei modelled which enable the motor approach towards either the yellow or green landmark. The agent detects the landmark through proximal and distal sensors. These proximal signals (Y-proximal and G-proximal) represent the US (US_y and US_g) processed by the dmPFC and generated by the yellow green landmarks respectively. These feed into the corresponding core units that enable motor control to the respective yellow or green landmarks. The distal signals G-distal and Y-distal of both landmarks which assume the role of the CS (Y_{CS} or G_{CS} from the yellow and green landmark respectively) are processed by the excitatory dmPFC projections to both neural core units. The G-distal (G_{CS}) signal activates the core-g and core-y units through weighted ρ_{gg} and ρ_{gy} synapses while the Y-distal (Y_{CS}) signal activates the core-g and core-y units through weighted ρ_{yg} and ρ_{yy} synapses respectively. These excitatory afferents are modulated by DA released from the VTA. The core uses the

distal and proximal signals to enable motor activity as has been described in the Fig. 4.1C.

The shell is also innervated by cortical inputs from the orbitofrontal region (OFC) of the PFC. The OFC maintains persistent activity triggered from visualising either of the landmarks for a set period or until a reward is obtained. Therefore, this activity goes beyond the US if omitted and can be used to generate extended tonic DA activity such that LTD is also extended. The g-distal and y-distal signals from the green and yellow landmarks respectively are processed by the OFC as working memory units which generate persistent activity Y_{PA} and G_{PA} to the shell through plastic ω_g and ω_y synapses respectively. They maintain activity for a set period if their activity reaches a threshold value.

Activation of the shell by the persistent OFC inputs results in the inhibition of the mVP. The mVP actively inhibits the VTA and the MD which release DA and projects back to the PFC respectively. The distal signals to the shell are capable of activating the shell and which in turn dis-inhibits the VTA and MD via the mVP. By dis-inhibiting the MD and VTA, the shell can indirectly influence the ability of the core to enable motor activity and the VTA neurons to release DA respectively. This means that shell activation by the distal signals can influence motor drive as well as DA release.

Information flow and weight change during acquisition in an idealised food seeking run is described next.

4.3 Information Flow and Plasticity in the NAc During Acquisition

An ideal cartoon of the reward seeking task during acquisition is described here with the intention of demonstrating information flow through the circuit. A real simulation run will be shown once the complete circuit has been established.

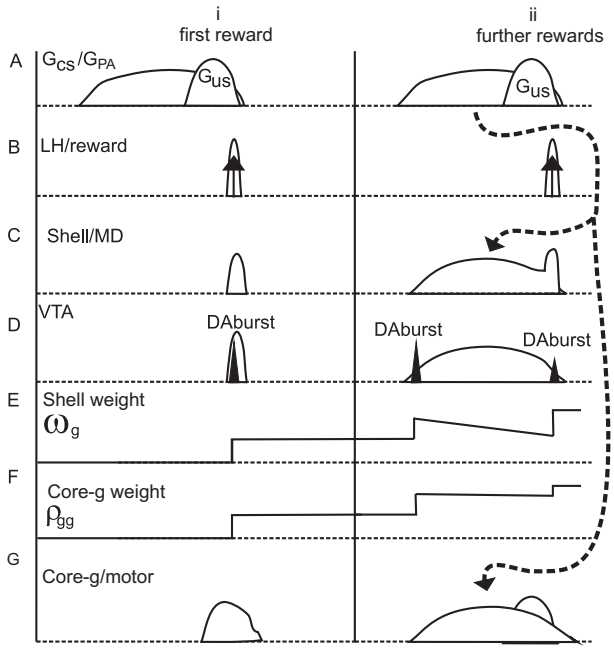

Figure 4.3: A cartoon of information development during the acquisition at two trials labelled i and ii against time (t). A) G_{CS} and G_{PA} represent signal generated from the green landmark as the agent approaches the landmark. These signals feed into the prefrontal and orbitofrontal cortex. B) LH activated when the agent receives the reward. C) Shell activity development which also illustrates MD dis-inhibition. D) VTA activity showing the two activity states. A DA burst is produced when the reward is obtained and eventually shifts to the CS onset. The DA burst occurring in event of reward receipt slowly decreases. E) ω_g, the shell weight development for the plastic synapses relevant to the cortical inputs which are activated by signals from the green landmark. F) ρ_{gg}, the weight development for plastic synapse signalling the green landmark projecting to the core-g unit. G) Core-g unit activity. (Thompson et al., 2009)

At the beginning of the run, the naïve agent wanders around the environment in which are yellow and green landmarks. The distal (X-distal) and proximal (X-proximal) signals generated by either the yellow (X=Y) or green (X=G) landmark (X) are bandpass filtered to represent the CS (X_{CS}) and US (X_{US}) signals respectively. Filters are used to simulate the responses observed by

sensors systems.

$$X_{US}(t) = h_{LP}(t) * X\text{-}proximal(t) \quad (4.1)$$

$$X_{CS}(t) = h_{LP}(t) * X\text{-}distal(t) \quad (4.2)$$

These signals are processed by the dmPFC which projects to the individual core-x units.

The distal signal projects via weighted plastic inputs to both the shell and the core. It is bandpass filtered and activity is maintained for a set period due to the OFC processing according to Eq.3.14. $u_j(t)$ is represented as X_{PA} and corresponds to persistent activity (Fig. 3.4) occurring in the input neuron from the yellow ($X = Y$) or green ($X = G$) landmarks.

Information flow and acquisition as the agent approaches the green landmark is shown in Fig. 4.3A. When in close proximity to a landmark, the proximal signal (X_{US}) triggers the agent's motor towards the center of the landmark X. In addition, if the agent comes in contact with the food reward in the green landmark, the LH becomes active according to Eq.3.1 (Fig. 4.3A i and ii). It is a bandpass filtered signal of the food *reward* signal.

The information processed by the LH, OFC and PFC summate onto the corresponding shell, core-g and core-y units according to equations 3.19 and 3.22 (Fig. 4.3A iii and vii).

$$shell(t) = LH(t) + (G_{PA}(t) \cdot \omega_g(t)) + (Y_{PA}(t) \cdot \omega_y(t)) \quad (4.3)$$

$$\begin{aligned} core\text{-}g(t) &= G_{US}(t) + (Y_{CS}(t) \cdot \rho_{yg}(t)) + (G_{CS}(t) \cdot \rho_{gg}(t)) \\ &- \lambda \cdot core\text{-}y(t) \end{aligned} \quad (4.4)$$

$$\begin{aligned} core\text{-}y(t) &= Y_{US}(t) + (Y_{CS}(t) \cdot \rho_{yy}(t)) + (G_{CS}(t) \cdot \rho_{gy}(t)) \\ &- \lambda \cdot core\text{-}g(t) \end{aligned} \quad (4.5)$$

The X_{CS} and X_{PA} facilitate the core-X units and the shell through weighted synapses ρ_x and ω_x respectively. These are associated with the NAc units and are influenced by landmark X. Note that the activity in the core gates the

action subsystem or attraction behaviour towards the centre of the landmark. By implementing reciprocal inhibition, the strongest core activity performs a "*winner take all*" process by inhibiting other core units via λ (Klopf et al., 1993; Suri and Schultz, 1999). Contact with the food reward enables the LH to produce an excitatory glutamatergic activity on the VTA (Fig. 4.3A iv).

$$VTA(t) = \kappa \cdot LH(t) \tag{4.6}$$

This results in a fast spiking DA burst defined by the VTA processed through a highpass filter with a strength χ_{burst} according to the Eq.3.9.

LTP requires both pre and post-synaptic activity as well as D1 receptor activation (Reynolds and Wickens, 2002) which is obtained via the burst spiking of DA neurons. Therefore LTP is modelled in the shell and core as follows:

$$\omega_X \leftarrow \omega_X + \mu_{shell}(X_{PA} \cdot shell' \cdot burst \cdot (limit - \omega)) \tag{4.7}$$

$$\rho_X \leftarrow \rho_X + \mu_{core}(X_{CS} \cdot core\text{-}X' \cdot burst \cdot (limit - \rho)) \tag{4.8}$$

Thus the DA burst enables the plastic weights ρ_x of the core-X units and ω_x of the shell to increase (LTP) via three factor learning (Fig. 4.3A v and vi).

4.4 The Simple Reward Seeking Experiment

This section demonstrates how the NAc plays a role in mediating reward seeking behaviour in a simple scenario task. The environment is setup as described in the previous section Fig. 4.1D. There are two landmarks, a yellow and a green landmark situated opposite each other by the central left and right walls of the environment. The reward is embedded in the green landmark. The agent starts at a starting point located North of the centre line in the environment and explores the environment in a straight trajectory

until it either comes in contact with the walls and navigates along them or it gets close to a landmark and elicits its pre-wired reflex towards the centre of the landmark. Fig. 4.4 shows the agents footstep during A, the first half of the full simulation run and B the second half of the simulation run when the agent has learned to associate the green landmark with the reward. During the first half of the simulation run between time steps 0 to 25000 (Fig. 4.4A), the agent can be seen to navigate along the walls of the environment until it comes in contact with a landmark and elicits an attraction behaviour towards the centre of the landmark. If the agent makes contact with the yellow landmark, it continues in a straight trajectory across the environment or along the walls of the environment until it comes in contact with the reward.

Once the agent comes in contact with the landmark containing the reward, it is repositioned at a random orientation and location on the midline equidistant to the two landmarks and the agent navigates again until it finds the reward. As learning progresses, the agent demonstrates a biased conditioned response (CR) towards the green landmark containing the reward Fig. 4.4B. It can be seen from the density of the agents trajectory towards the green landmark from the central line that the agent learns to approach the green landmark from a distance.

Detailed signal traces of the CS, LH, VTA, DA bursts and weight development occurring between the time steps 15000 to 35000 and 40000 and 45000 are illustrated in Fig. 4.5. As the agent approaches the green landmark from a distance, the distal signal generated by the green landmark (Y_{CS}) becomes high. When the agent finds the reward located in the green landmark, the LH and a VTA DA burst become active which correlate with the high G_{CS} activity so that its corresponding plastic weight ω_g increases (highlighted regions i, iii and v).

There are five significant events highlighted and numbered i to v which show how the DA burst at the US (i) onset decreases in amplitude (iii and v) and increases at the CS onset (ii and iv). The DA bursts generated at the CS

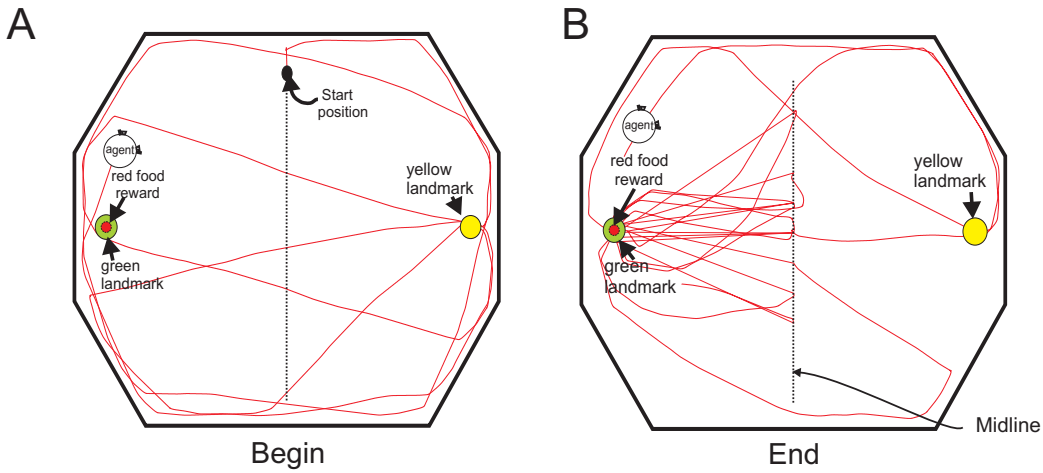

Figure 4.4: The agent trajectory during the first (begin) and last (end) half of the simulation run over a period of 50000 time steps. A) The first half of the simulation run, the agent wanders along the walls of the environment and performs a reflex reaction towards the centre of any landmark. The agent begins from the location indicated start position. During the first half of the simulation run, the agent comes in contact with the reward only once. B) The second half of the simulation run the agent's learned reaction is towards the green landmark with the reward. During the second half of the simulation run, the agent makes contact with the reward seven times. When the agent makes contact with the reward, the agent is repositioned at a random position in the midline. Simulation parameters are presented in appendix C in table C.2.

onset can be used to develop further associations between more than one CS.

The models performance in the simple reward seeking experiment is observed in ten simulation runs, each conducted over a duration of 100,000 time steps. Fig. 4.6 shows the average number of contacts to the green and yellow landmark made after 10,000 time steps over the full duration of the run. It can be seen that the agent makes significantly more approaches to the green landmark that to the yellow landmark over the course of the simulation run. These results show that the agent learns over time, to approach the green landmark which contains the reward.

The performance of the model in this reward seeking experiment can be

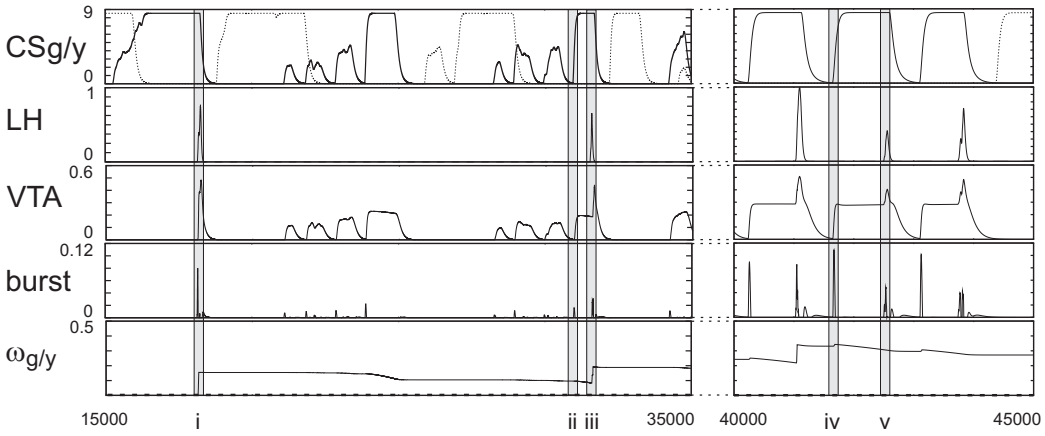

Figure 4.5: Detailed signal traces of the simulation run for the periods of 15000 to 35000 and 40000 and 45000 time steps. Traces include the distal signals for the yellow and green landmarks (CS), The LH, the VTA, burst and weight development ($\omega_{g/y}$). Five significant events labelled i to v have been highlighted to show how the DA burst generated at the US event decrease and increase at the CS onset. Simulation parameters are presented in appendix C in table C.2.

compared against results obtained by real rats. In the in vivo experiments two visual stimuli were presented which were either followed by a reward or no reward delivery. Although the experimental setup in the in vivo and simulated experiments are different, both experiments contain rewards and stimuli that are associated with the reward (CS+) and stimuli that are not (CS-). The agents must demonstrate that they have learned to elicit a CR in response to the CS+. The following sections briefly introduces the animal experiment conducted by Parkinson et al. (2000). The model will be tested in the simulation experiments so that the results produced will be compared against the empirical results generated by Parkinson et al. (2000).

4.5 Comparison Against Empirical Data

Parkinson et al. (2000) studied the performance of rats that were subject to shell and core lesions in Pavlovian approach behaviour. During Pavlovian

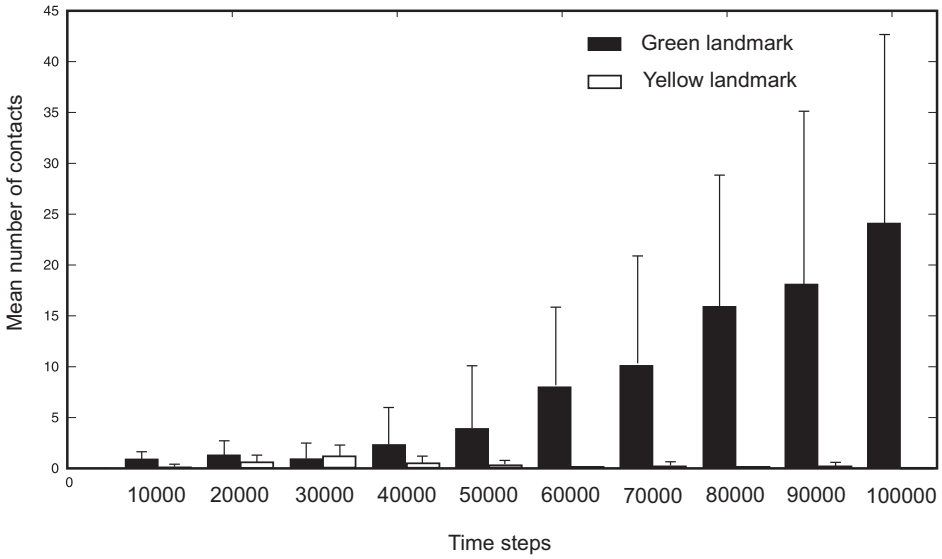

Figure 4.6: Simulation results showing the mean number of contacts to the green (shaded boxes) and yellow landmarks (clear boxes) as a function of 10,000 time steps incrementing to 100,000. Bars indicate the standard deviation of 10 runs. Simulation parameters are presented in appendix C in table C.2.

approach behaviour, a visual stimulus is presented after which a reward is delivered. After training, the animal generates a conditioned response (CR) approach to the visual stimulus (CS) before the food is presented.

The apparatus implemented was a testing chamber which contained a visual display unit (VDU) that presented the visual stimuli. The chamber also contained a food hopper in which sucrose pellets (reward) were delivered. There were floor pads that detected the animals location in the chamber and identified when the animal was at a location equidistant to the two stimuli locations. The VDU displayed the visual stimuli and the rats responses were measured through the use of touch sensitive floor pads and screens. The animals were trained to associate a stimuli with sucrose pellets as the reward. This was identified as the CS+. The CS- was the stimulus presented that was never followed by the reward. During the tests, CSs were presented and the animals responses were recorded. All of the animals were divided into

groups some of which were subjected to excitotoxic lesions of their core and shell regions. Each groups performance in response to the CS was recorded. In these studies the animals performance in the tasks were affected by core rather than shell lesions.

A behavioural experiment is simulated and used to test the model's performance in a simple reward seeking experiment. In this experiment, a reward is embedded in a landmark and the agent learns a simple association between the reward and the landmark. In the simulation experiments, the model will be subject to simulated shell and core lesions and its performance will be compared against the in vivo results obtained by Parkinson et al. (2000).

The following section illustrates how the model performs when subject to either shell and core lesions. The results are presented in such a way that the model's performance can be compared against the results generated by Parkinson et al. (2000).

4.5.1 The Simulated Lesion Experiments

The performance of the model subjected to either core or shell lesions are presented in Fig. 4.7A and Fig. 4.8A respectively. These are presented along with the adapted results from the core (Fig. 4.7B) and shell (Fig. 4.8B) lesion experiments conducted by Parkinson et al. (2000). The results in Fig. 4.7A and Fig. 4.8A show the mean response contacts to the CS+ and CS- over ten simulation runs, while the results in Fig. 4.7B and Fig. 4.8B illustrate the results adapted from Parkinson et al. (2000) of the mean approaches to the CS over blocks of stimulus presentations. The core and shell lesions are simulated by reducing but not completely eliminating weight connectivity of the core motor-enable pathways or the shell-VTA and shell-VP pathways respectively.

The results illustrated in Fig. 4.7A show that the performance of the model with core lesions was impaired compared with the model that was not lesioned. The lesioned agent's approach to the CS+ was significantly reduced

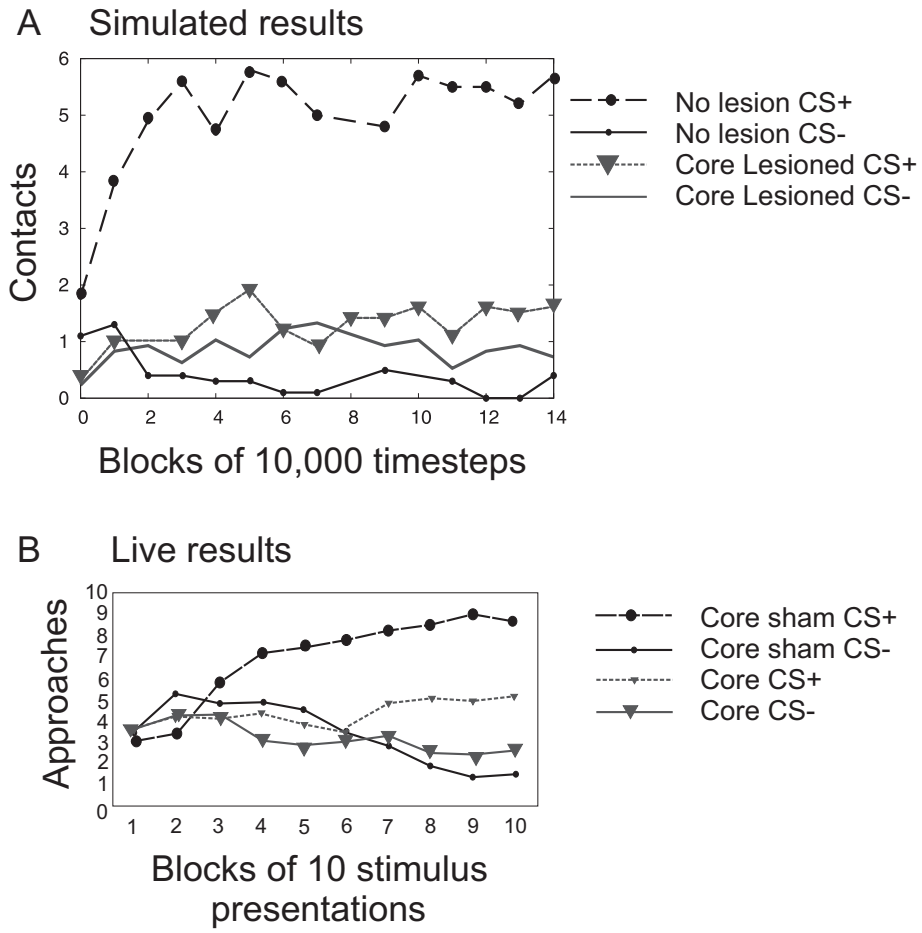

Figure 4.7: A) The model's performance in the behavioural food seeking task when subjected to core lesions compared with the model with no lesions. B) Adapted results from Parkinson et al. (2000) showing the acquisition of autoshaping behaviour after lesions to the core and sham lesioned rats. Simulation parameters are presented in appendix C in table C.2.

however the agent was still capable of approaching the either of the landmarks when it was in close proximity to the landmarks. The model achieves this because the US signal is still strong enough to project through the core units and enable motor behaviour towards the landmarks. If the core units were completely lesioned, the agent would not demonstrate a response to the US. The current model might achieve an ability to respond to the US if the network via the dorsal striatum to the basal ganglia were connected to the

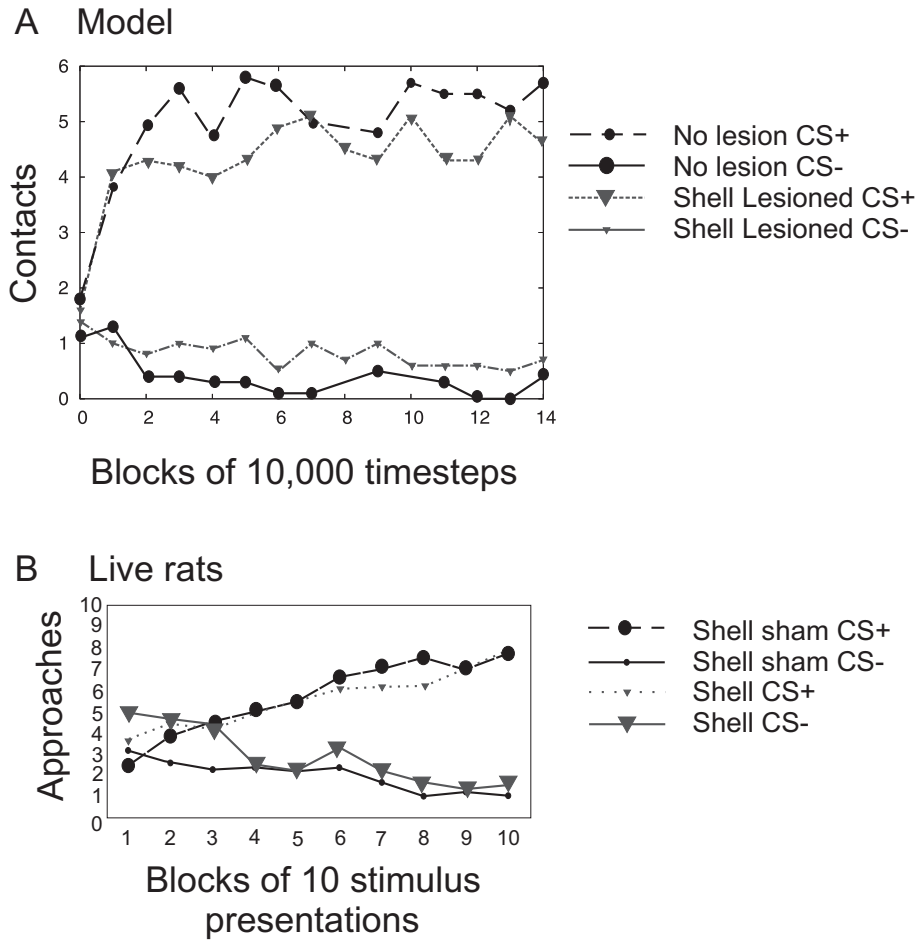

Figure 4.8: A) The model's performance in the behavioural food seeking task when subjected to shell lesions compared with the model with no lesions . B) Adapted results from Parkinson et al. (2000) showing the acquisition of autoshaping behaviour after lesions to the shell and sham lesioned rats. Simulation parameters are presented in appendix C in table C.2.

current model. This would require significant modification that is beyond the scope of this work. The acquired CS signals on the other hand are not strong enough to enable motor activity. Therefore, the performance of the core lesioned agent is severely impaired compared to the performance of the non-lesioned agent.

Unlike the results illustrated in Fig. 4.7A, Fig. 4.8A shows that the perfor-

mance of the model when simulated with lesions to the shell were not impaired and performed similarly with the non-lesioned model. This is because acquisition in the core units were sufficient to mediate approach behavior in response to the CS+. The shell units influence approach behavior when the US is omitted and the CS no longer precedes the US. However, these conditions were not tested for in these experiments. Similar to the non-lesioned agent's approach to the CSs, the lesioned agent's approach to the CS+ was significantly greater than the approach towards the CS-.

The results illustrated in this section are similar to the results generated by Parkinson et al. (2000) represented in Fig. 4.7B and Fig. 4.8B. In these in vivo experiments, rats with lesions to the NAc core were severely impaired during acquisition. These rats did not show any significant discriminated approach to the CS. The results generated in the simulation resembled the results produced by the rats with shell lesions. Both the computational model and the rats did not demonstrate any significant impairment during acquisition. These results are the most recent generated from this study. They show how the modelled shell and core units function similar to the shell and core regions of the biological agent.

It is also essential that the agent is capable of distinguishing between relevant and irrelevant stimuli and adjusting behaviour as contingencies change in the environment. A reversal learning experiment is performed in the following section and utilised to demonstrate how the NAc plays a role in goal directed behavioural flexibility by the inhibition of rather than elimination of acquired associations.

4.6 The Reversal Learning Task

Although the core circuit is sufficient for acquisition described so far, the influence from the shell in facilitating behaviour is necessary when the reward is omitted from the green landmark and the agent must inhibit behaviour towards the green landmark which no longer contains the food reward. The

model adapts as contingencies change, not by the standard method of eliminating the originally acquired associations which does not account for rapid reacquisition, but by disabling the gating of irrelevant action subsystems. The role of the shell in implementing this mechanism, by indirectly influencing the core activities will be shown here.

Reversal learning experiments conducted by Birrell and Brown (2000) and later by Egerton et al. (2005) have been simulated so as to test the computational model. In the in vivo experiments, rats are placed in an environment which contains two digging holes both emitting distinct odors and one of which contains food pellets. The rats are required to associate an odor with the reward and learn to go directly to the digging hole with the odor associated with the reward. After the rat has demonstrated acquisition for the odor coupled with the reward while completely ignoring the opposite hole, the contingency is reversed so that the food pellet is now placed in the second hole which originally lacked the reward. The rats need to learn to inhibit their behaviours towards the hole which originally contained the reward and learn to associate the second hole with the reward.

4.6.1 The Reversal Learning Simulated Environment

In this octagonal environment are two landmarks coloured yellow and green and an agent which explores the environment for rewards which are embedded inside the landmark indicated by the red disk. The agent is required to learn an association between the landmark and the reward disk and to approach the landmark containing the reward from a distance. It can only detect the reward when it makes direct contact with it. Associations are acquired between the distal signal (CS) and proximal signal (US) from the landmark containing the reward.

Once the agent demonstrates that it has learned to approach the landmark from a distance, the reward is no longer placed in the green landmark but instead is now placed in the yellow landmark. The agent now has to inhibit

behaviour towards the green landmark and learn to associate the yellow landmark with the reward. In the following section, a computational circuitry necessary to perform acquisition and reversal respectively is developed.

While the core enables motor activity to elicit behaviours in response to the reward predictive stimulus, the shell indirectly facilitates the inputs to the core to drive the acquired behaviours via the shell-mVP-MD pathway. The reversal learning scenario during which the agent demonstrates behavioural flexibility is described in the following section.

4.6.2 Information Flow and Plasticity in the NAc During Reversal

Reversal learning begins when the reward is omitted from the green landmark and placed in the yellow landmark. Fig. 4.9 shows information flow during reversal learning when the agent approaches the green landmark after the reward has been omitted. The agent having learned to associate the green landmark with the reward, exhibits behaviour towards the green landmark (Fig. 4.9A). At this stage there is no LH activity due to the absence of a reward (Fig. 4.9B. The shell which becomes active due to the high weight (ω_g) dis-inhibits both the VTA and MD (Fig. 4.9 panels C and D). Consequently, to reflect the dis-inhibition from the mVP (Eq.3.20), Eq. 4.6 needs to be updated based on the excitatory, inhibitory and dis-inhibitory influences from the LH, shell and shell-mVP pathways respectively. The equation representing the VTA is updated according to Eq.3.7.

The absence of LH activity and the dis-inhibition of the VTA by the shell generates an increase in VTA activity proportional to the shell dis-inhibition only (Fig. 4.9 panels C and D). Thus the shell activation results in the dis-inhibition of the VTA and MD through the shell-mVP pathway (Eq.3.21).

The VTA dis-inhibition generates an increase in the population of the tonically active DA neurons detected as lowpass filtered VTA activity (Eq.3.10). This tonic DA activity (Fig. 4.9D) enables LTD to occur proportional to

presynaptic ($u_{X\text{-}distal}$) influences from the landmark X. A tonic activity which enables LTD in the absence of a DA burst, produces a resultant weight decrease in the NAc.

$$\begin{aligned} \rho_X(t) \leftarrow \rho_X(t) \quad &+ \quad \mu_{core}(X_{CS}(t) \cdot core\text{-}X(t)' \cdot burst(t) \cdot (limit - \rho_x(t))) \\ &- \quad \epsilon_{core}(u_{X\text{-}distal}(t) \cdot tonic(t)) \qquad (4.9) \\ \omega_X(t) \leftarrow \omega_X(t) \quad &+ \quad \mu_{shell}(X_{PA}(t) \cdot shell(t)' \cdot burst(t) \cdot (limit - \omega_x(t))) \\ &- \quad \epsilon_{shell}(u_{X\text{-}distal}(t) \cdot tonic(t)) \qquad (4.10) \end{aligned}$$

Here $\epsilon_{shell} \gg \epsilon_{core}$. This means that LTD in the shell occurs significantly more quickly than in the core (Fig. 4.9 panels E and F). A stronger LTD in the shell than in the core produces a swift decay of the shell weights to baseline (Fig. 4.9 panel E) until persistent activity no longer drives the shell. Slower LTD in the core ensures that learned weights (ρ_{gg}) are maintained such that the agents capacity to approach the landmark from a distance is not eliminated although the agent is required to inhibit approach behaviour towards the currently irrelevant landmark. The shell's ability to dis-inhibit the MD through the shell-mVP-MD pathway is diminished resulting in a decreased MD activity and an overall decrement in the cortical facilitation of the core unit (Fig. 4.9 panels C and G).

The cortical projections into the core are influenced by the MD innervations to represent the CS (X_{CS}) signal obtained from landmark X is updated:

$$X_{CS} = u_{X\text{-}distal}(t) + \theta_{MD} MD(t) \qquad (4.11)$$

Therefore, the shell (indirectly via the mVP-MD pathway) reduces the PFC activation on the core units such that the approach behaviour towards the irrelevant landmark is minimized.

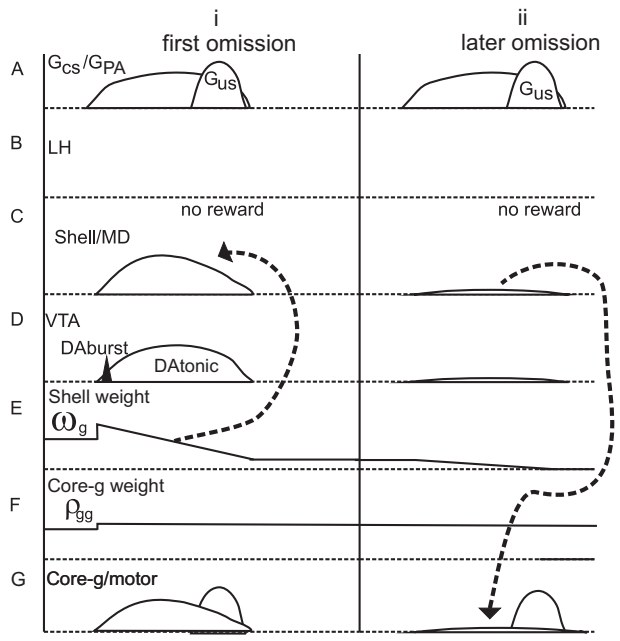

Figure 4.9: A cartoon of information development during reversal at two trials labelled i and ii against time (t). The reward is omitted from the green landmark. A) G_{CS} and G_{PA} represent signal generated from the green landmark as the agent approaches the landmark. These signals feed into the prefrontal and orbitofrontal cortex. B) No LH activation due to absent reward. C) Shell activity development which also dis-inhibits the MD. Shell activity decreases as LTD dominates in the absence of DA bursts. D) VTA activity showing the two activity states. A DA burst is produced during the CS onset. No further DA bursts are produced. E) ω_g The shell weight development occurring as a resultant decrease for the plastic synapses relevant to the cortical inputs which are activated by signals from the green landmark. F) ρ_{gg} The weight development for plastic synapse signalling the green landmark projecting to the core-g unit. G) Core-g unit activity.

4.6.3 Simulating Reversal Learning

The agent begins from the starting point Fig. 4.1A equidistant to both landmarks. Fig. 4.10 shows results of detailed information flow and weight development in the circuitry from the first acquisition to the first reversal occurring between time steps of 2000 to 45000. The agent wanders around the environment until it encounters a landmark during which it produces a curiosity

reaction towards the centre of the landmark. The box labelled I is the region which has been magnified in Fig. 4.11. It focusses on the highlighted regions numbered i, ii and iii. Contact with the reward for the first time is highlighted in the grey region of Fig. 4.10 and Fig. 4.11 labelled i. During this event, the OFC activity produced by the signals from the green landmark is high and coincides with the LH activity generated by obtaining the reward in the green landmark. This causes spiking VTA activity and resultant phasic levels of DA and LTP in the NAc. However, VTA DA burst is not only generated at LH activation but also via the shell-mVP-VTA pathway. This is responsible for the VTA burst at the CS onset. In other words, once the reward becomes predictable, the DA bursts start occurring earlier at the onset of the cue that predicts the reward. In this case, the CS that predicts the reward is represented by the distal signals which also trigger OFC activity onset.

LTP on the ω_g synapse enables increased OFC activity in the shell and stronger dis-inhibition of the VTA. This means that as the weight increases, an amplified activity in the shell enables the spiking activity of DA neurons to occur more regularly. In this way the DA bursts occur during the CS onset. The arrow in the highlighted grey region numbered ii shows how the DA burst at the CS event increases in magnitude as the shell activity increases. Because the shell's ability to inhibit the mVP is capped at a minimum value (VP_{min} from Eq.3.20), there comes a point when the increasing shell activity starts to inhibit the VTA DA neuron more strongly than both the LH influence and its dis-inhibition on the DA neurons (time steps between 10000 and 20000). This is established by the direct shell-VTA pathway and its effect can be observed in the decreasing burst spiking DA activity occurring at the US onset as shown by the arrow in the highlighted region numbered iii. Eventually, the DA bursting activity at the US onset decreases to baseline.

The agent demonstrates that it has acquired an association between the green landmark and the reward when it makes ten consecutive contacts with the reward. After this, the reward is moved from the green landmark to the yellow landmark. The arrow labeled reversal denotes that the contingency

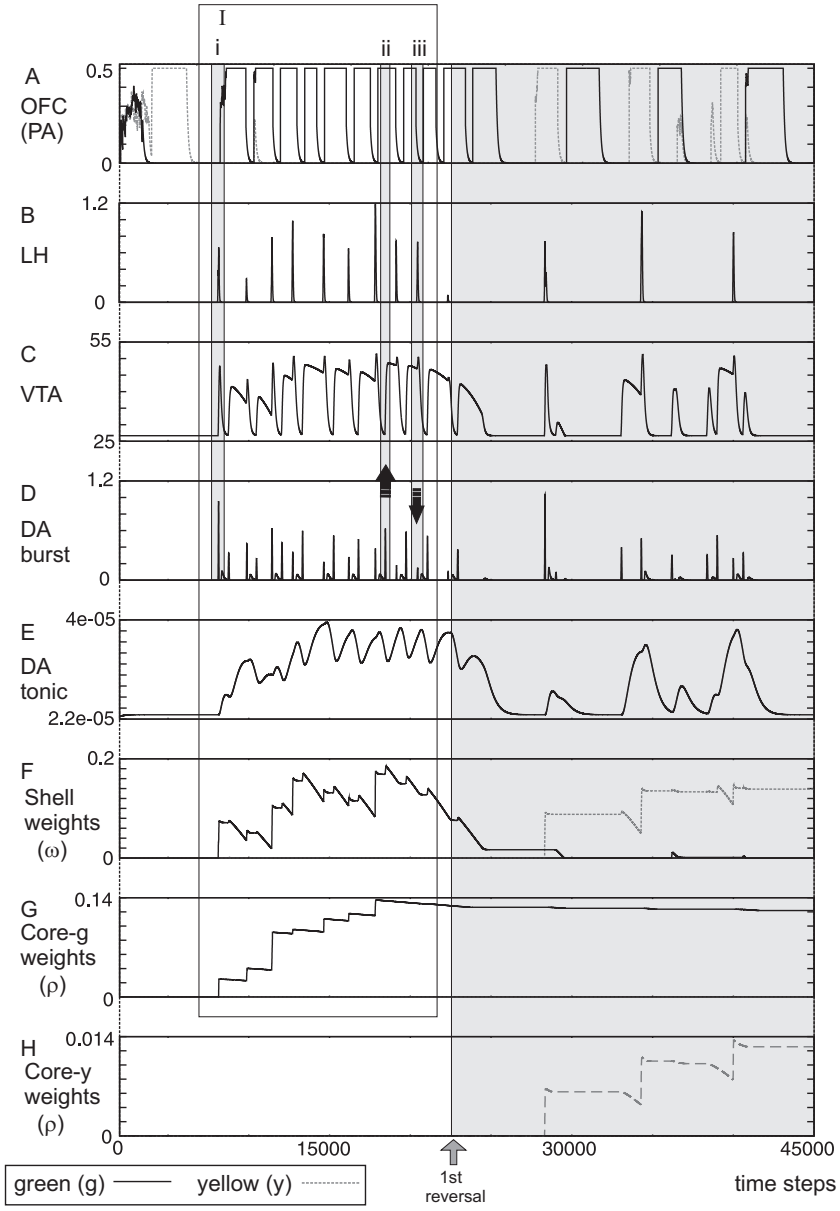

Figure 4.10: The activity of the A) OFC inputs B) LH C) VTA D) Burst E) Tonic F) Shell weights G) Core-g weights F) Core-y weights. The box I contains the highlighted regions numbered i, ii and iii. I is magnified to show the DA bursts occurring in the highlighted regions. Simulation parameters are presented in appendix C in table C.2.

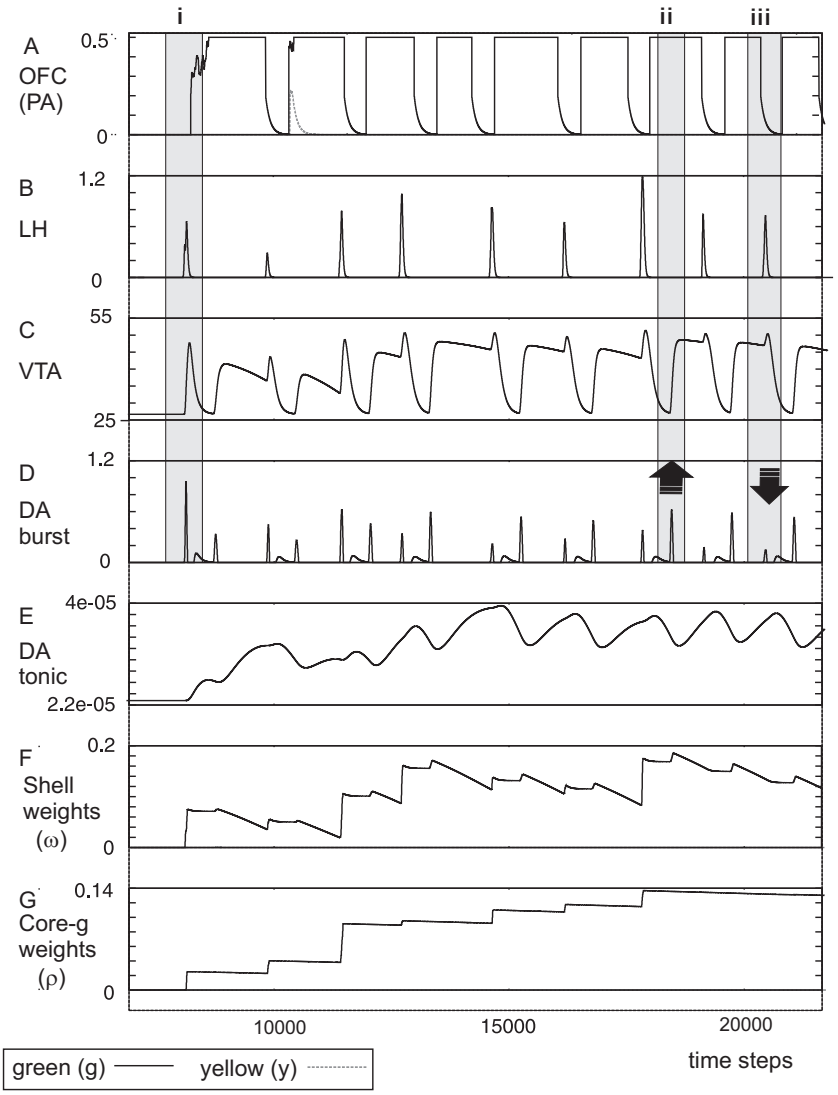

Figure 4.11: The magnification of the region labelled I in the detailed signal trace showing the highlighted region numbered i ii and iii. i indicates the first DA burst at the US event. While the upward and downward arrows in the highlighted regions ii and iii respectively indicate increasing and decreasing DA burst at the US and CS events. The activity of the A) OFC inputs B) LH C) VTA D) Burst E) Tonic F) Shell weights G) Core-g weights Simulation parameters are presented in appendix C in table C.2.

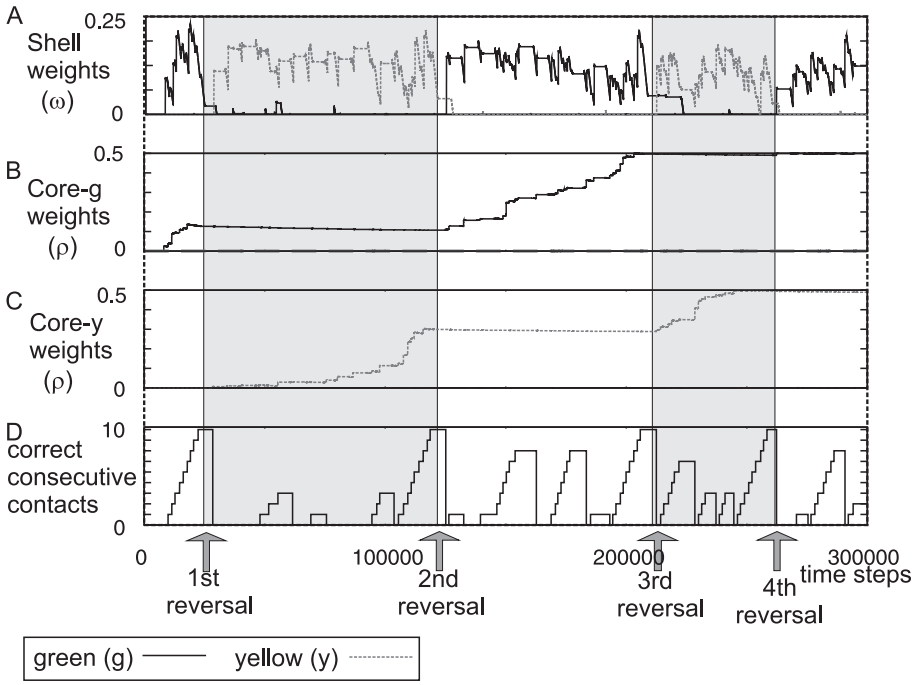

Figure 4.12: The activity of the A) Shell weights B) Core-g weights C) Core-y weights D) Number of correct consecutive contacts during four reversals. Simulation parameters are presented in appendix C in table C.2.

has changed and reversal learning commences. The OFC activity generated by the green landmark is observed to persist longer than previous activations. This is because the OFC enables persistent activity for a set period or until the reward is obtained. The OFC activates the shell which in turn disinhibits the VTA activity to produce tonic DA levels that enable LTD to occur on the synapses in the shell that are currently active. The dotted lines in Fig. 4.10A correspond to OFC activation by the signals from the yellow landmark. Eventual contact with the reward in this landmark generates LTP on the ω_y synapses and the whole process repeats itself but this time for an association between the yellow landmark and the reward.

The shell and both core units weight development for a simulation run over a period of 300,000 time steps is shown in Fig. 4.12. Here the contingency is reversed four times. It can be seen that while the shell weights increase

and decrease rather quickly, the core weights increase quickly but decrease at a much slower rate. Learned behaviours are maintained in the core and reversal learning is achieved instead via the shell which updates the relevant information and mediates the cortical activity to the core. It can be seen in Fig. 4.12, that the duration between the third and fourth reversal is smaller than the duration of the first and second reversal. This shows how the model is capable of performing later reversals more quickly because associations do not have to be relearned. A clip showing the agent performing reversal learning can be viewed at http://isg.elec.gla.ac.uk/maria/reversal/.

The performance of the model can be compared against the performance of animals tested in serial reversal learning experiments conducted by Bushnell and Stanton (1991) and Watson et al. (2006). These experiments are briefly summarized.

4.7 The Model's Performance against in Vivo Serial Reversal Learning Experiments

The serial reversal learning experiments carried out by Bushnell and Stanton (1991) were conducted on rats in an apparatus that was composed of two retractable levers, a cue light and a food cup in which food was delivered. The rats were required to press one of two levers for food rewards. The lever that resulted and did not result in reward delivery are referred to as the CS+ and CS- respectively. Each reversal occurred when a criterion was met that was determined by the discrimination ratio (DR). The DR was defined as follows: DR = frequency of correct / (frequency of correct + frequency of wrong). During reversal the CS+ became the CS- and vice versa so that they respectively produced an opposite result to that which they originally predicted. A criterion was met when the discrimination ratio (DR) reached or exceeded a value of 0.9 for two consecutive ten trial blocks.

In the serial spatial reversal learning experiments conducted by Watson et al.

(2006), rats were tested using a T-maze. The T-maze consisted of three arms of equal length which made up the start, left and right arms. There were photo beams which detected when the rats had approached any one of the arms. During the test, the rats were made to begin in the start arm. Shortly after, they were allowed access to the left and right arm one of which was allocated as the correct arm. Rats were rewarded when they entered the correct arm. After a certain period, the reversal session occurred and the rats were rewarded on entering the opposite and previously unrewarded arm.

The model's performance in the serial reversal food seeking task was tested over ten simulation runs which lasted over a maximum duration of 500,000 time steps. The average number of total contacts made per reversal for one original discrimination and five reversals over ten simulation runs is illustrated in Fig. 4.13A. These results are compared against the adapted results from Bushnell and Stanton (1991) in Fig. 4.13B which show the mean trials to criterion per reversal for one original discrimination and five reversals. Similar to the serial reversal experiments performed by Bushnell and Stanton (1991) the criterion for reversal were also determined by the DR. However, for suitability, the DR value for which the criterion was to be met was set to 0.7 over 20 contacts with either landmarks. The results from the in vivo experiment and the simulated runs show that on average, the contacts or number of trials required to meet criterion were smallest for the first acquisition but were at maximum values during the first and second reversals. These values decreased as the reversals were repeated.

The average errors made per reversal are presented in Fig. 4.14 and can be compared against the errors made by real rats in the serial reversal learning experiments carried out by Watson et al. (2006). The broken line in Fig. 4.14 illustrates an adapted result from Watson et al. (2006) and represents the mean total number of errors made across one original discrimination and five reversals for rats that were 26 postnatal days old. Both the simulated and real experiments showed that learning improved across reversals such that there were fewer number of errors made per reversal as the reversals were repeated. The reversal and serial reversal learning experiments presented

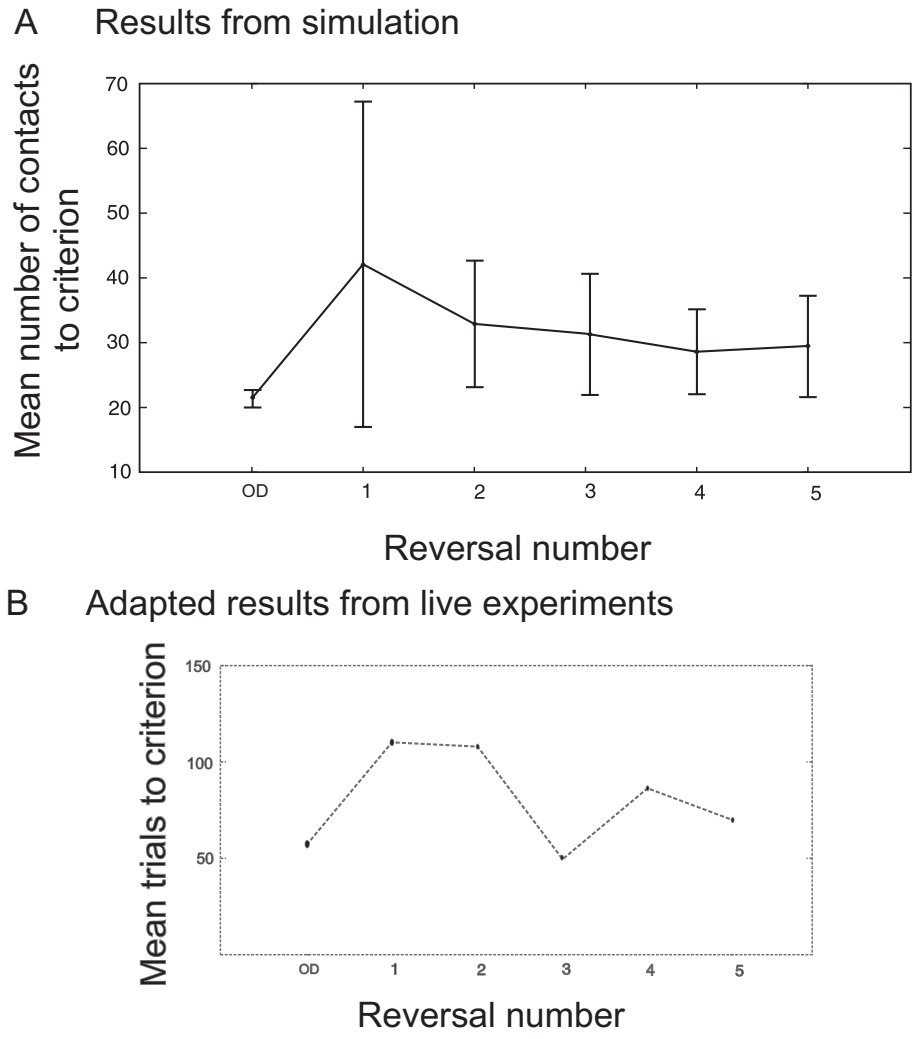

Figure 4.13: A) Serial reversal learning curve obtained from ten simulation runs showing the mean contacts to criterion across an original discrimination (OD) and five reversals as numbered. Bars indicate the average and standard deviation of ten runs which show the mean trials to criterion plotted as a function of reversal. B) Adapted serial reversal learning curve from Bushnell and Stanton (1991) Simulation parameters are presented in appendix C in table C.2.

here are similar to the experiments provided in (Thompson et al., 2009).

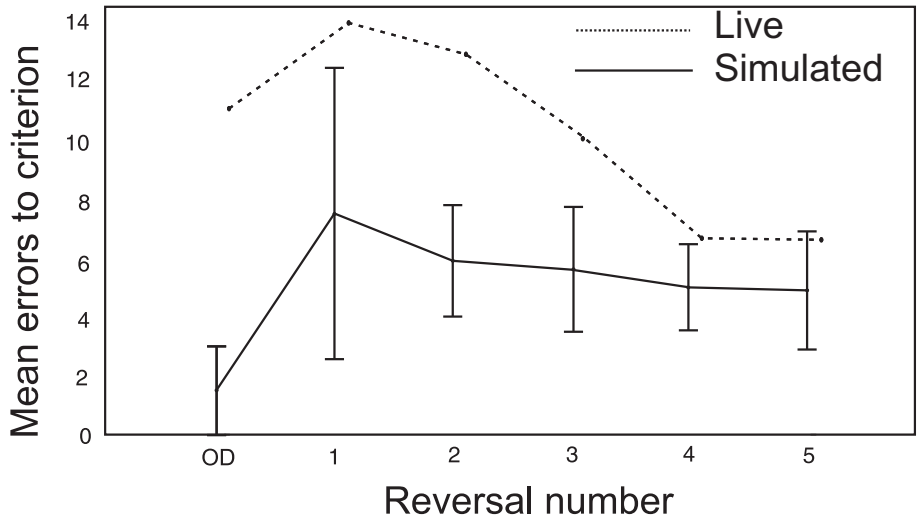

Figure 4.14: Serial reversal learning curve showing the mean total number errors made per reversal. Bars indicate the average and standard deviation of ten runs which show the mean trials to criterion plotted as a function of reversal. The broken line represents the adapted result from Watson et al. (2006) which illustrate the mean total number of errors made as a function acquisition and five reversals by rats that were 26 days old. Simulation parameters are presented in appendix C in table C.2.

4.8 Concluding Remarks

In this chapter the model developed in chapter 3 was tested in a variety of closed-loop behavioural reward seeking experiments. These experiments have been regarded as closed-loop experiments because the agents response in the environment strictly determine whether or not they obtain rewards. The closed-loop behavioural experiments were introduced by describing a scenario environment and demonstrating how the agent interacted with the environment and developed behaviour as a result. A simple reward seeking experiment was illustrated. The model's performance in this behavioural experiment subject to simulated core and shell lesions were associated with and compared against in vivo experiments conducted by Parkinson et al. (2000) which also involved rats with shell and core lesions. These comparison experiments are considered as closed-loop experiments because the approach

response determined if a reward was obtained.

In the third and final experiment, the agent had to discriminate between a CS1 and reward while a second CS (CS2) did not predict the reward. After this acquisition, the contingency was reversed such that the CS1 no longer predicted the reward and CS2 predicted the reward. This process was repeated a few times. The model's performance in this serial reversal learning experiment was compared against in vivo experiments conducted by Bushnell and Stanton (1991) and Watson et al. (2006). The model's performance was similar in some respects, to the results from the in vivo experiments. These comparisons can be used to indicate how the model may be compared to in vivo serial reversal learning experiments. However, more simulations are required to improve and substantiate the comparisons made at this stage.

This chapter shows that the model is capable of performing in a variety of closed-loop experiments. In addition when the model was subjected to simulated shell and core lesions it performed in a similar way to rats that had also been lesioned in their shell and core regions. The agent also performed in a similar way to rats in the serial reversal learning experiments. In the next chapter the model will be compared against some computational models that have also been intended to model animal learning.

Chapter 5

A Comparative Study of the Sub-cortical Limbic Model

This thesis has proposed a biologically motivated computational model of the sub-cortical nuclei of the limbic system which is capable of performing learning and reverse learning in reward based tasks. The model acquires associations using an extended correlation based differential Hebbian learning rule known as three factor Isotropic Sequence Order (ISO3) learning. The third-factor or modulatory signal is identified and modelled as phasic dopaminergic activity which in biology is associated with reward processing. When an unexpected reward is obtained, the phasic dopamine (DA) activity enables weight increase or long term potentiation (LTP). In order to demonstrate behavioural flexibility, the model uses a rise and not a pause of tonic dopaminergic activity to enable weight decrease. In addition, when rewards are omitted, acquired stimulus-response associations are attenuated rather than abolished. This is achieved by employing a feed-forward value switch that facilitates and attenuates sensor inputs accordingly.

The feed-forward pathway represents the pathway between the shell and the cortical projections to the core via the mediodorsal nucleus of the thalamus (MD). The computational model has been developed and tested in both classical conditioning and closed-loop behavioural experiments as described

in chapters 3 and 4. In this chapter, the importance of the feed-forward pathway in behavioural flexibility will be observed.

The model is represented as a modified actor-critic architecture whereby the circuitry surrounding the shell and the core represent the critic and actor respectively. The model as an actor-critic model, is made up of an additional feed-forward component. It mediates behavioural flexibility by facilitating and attenuating the sensor inputs. In the following sections a comparison will be made between three different versions of the computational model categorised according to the rate by which "*unlearning*" occurs in the actor i.e. the rate of LTD in the core, and whether or not there is a feed-forward connection between the critic (shell) and actor (core). The models compared are; a limbic circuitry equivalent of the classical actor-critic model, the actor-critic equivalent model developed in this thesis and a hybrid model of the previous two models. Each architecture's ability to account for rapid reacquisition will be observed in an open-loop experiment. The performance of the models will also be compared against each other in the serial reversal learning food seeking task. A comparison will be measured according to the number of reversals achieved over a fixed number of time steps and according to the number of errors produced during the simulation runs. The number of errors can be quantified by observing the number of wrong contacts made per reversal. The results obtained will show that by implementing a feed-forward switching mechanism, the overall performance of the model in demonstrating behavioural flexibility is significantly improved.

5.1 The Model as an Actor-Critic Model

Actor-critic methods comprise two separate structures namely, the actor and the critic. The actor stores information about state - action (or stimulus - response) associations and selects actions based on this information, while the critic as its name suggests, criticizes the actions selected by the actor (Sutton and Barto, 1998). In classical actor-critic architectures, the critic uses a

temporal difference (TD) method (Sutton and Barto, 1982, 1987, 1990) to calculate an error signal which is then used to train the actor. The TD error becomes positive when an unexpected reward is obtained. During reversal, when the reward is omitted, a negative error is produced which depletes the learned stimulus-action association. A positive and negative error signal respectively result in weight increase or decrease. Therefore the direction of weight change in the actor and critic can be positive (*learning*) or negative (*unlearning*).

These actor-critic architectures acquire associations between stimuli and actions that allow for respective actions to be executed. When rewards are omitted, these acquired associations become destroyed again. As mentioned previously, the depletion of learned associations seems to be a rather inefficient and biologically unrealistic mechanism, currently implemented by the classical actor-critic computational models. In addition, it does not account for animal behaviour such as rapid reacquisition (Pavlov, 1927; Napier et al., 1992) which suggests that learned behaviours are not simply eliminated (Rescorla, 2001) during omission as reviewed by both and Bouton (2002) and Rescorla (2001). The model implemented so far, proffers a more efficient way that is supported by biological systems. Rather than perform rapid unlearning of already learned associations, the model suppresses or disables stimulus response pathways such that they can be quickly reactivated when necessary. The biological pathway that corresponds this theory includes the mediodorsal nucleus of the thalamus (MD).

The model, represented as an actor-critic architecture, is illustrated in Fig. 1.3. The actor corresponds the core circuitry while the critic involves the shell and its connectivity to the DA neurons of the VTA. There are three major differences between the classical actor-critic model which implement the TD-error and the current computational model as follows:

1. The TD methods utilized by numerous actor-critic models to simulate DA activity are replaced by an easily decodable phasic and tonic DA activity. Here, the positive and negative values of the TD error repre-

sented by the DA neurons (Schultz, 1998) are encoded by a burst and tonic DA activity which respectively result in LTP and LTD.

2. Tonic activity which produces LTD occurs at different rates in the actor and the critic such that it is much slower in the actor. This means that the actor does not immediately unlearn actions when rewards are omitted.

3. Actions are disabled by the critic through a feed-forward switching mechanism referred to as the value switch. This switch acts on the sensor input of the actor.

In this chapter, the value switch is proposed to play a major role in demonstrating behavioural flexibility. The importance of this value switch is observed by testing the performance of three versions of the models against one another in the rapid reacquisition tests and the serial reversal behavioural experiments as conducted in the previous chapter. The three different versions of the model are considered according to the MD pathway and the rate of LTD in the actor. The mechanism by which dopamine signalling and the MD pathway (value switch) is utilised in the model will be addressed.

5.1.1 The Actor Critic Models

Limbic circuitry as an actor-critic model are implemented as three different versions namely, the full-LTD (fLTD) model, the partial-LTD (pLTD) model and the partial-LTD-MD-feedforward (pLTD-MD) model. The models are briefly described as follows:

The full-LTD (fLTD) model

The full-LTD (fLTD) model implements a strong LTD rate in the actor so that learned stimulus-action associations are unlearned very quickly and there is no feed-forward switch between the critic and the actor. The fLTD

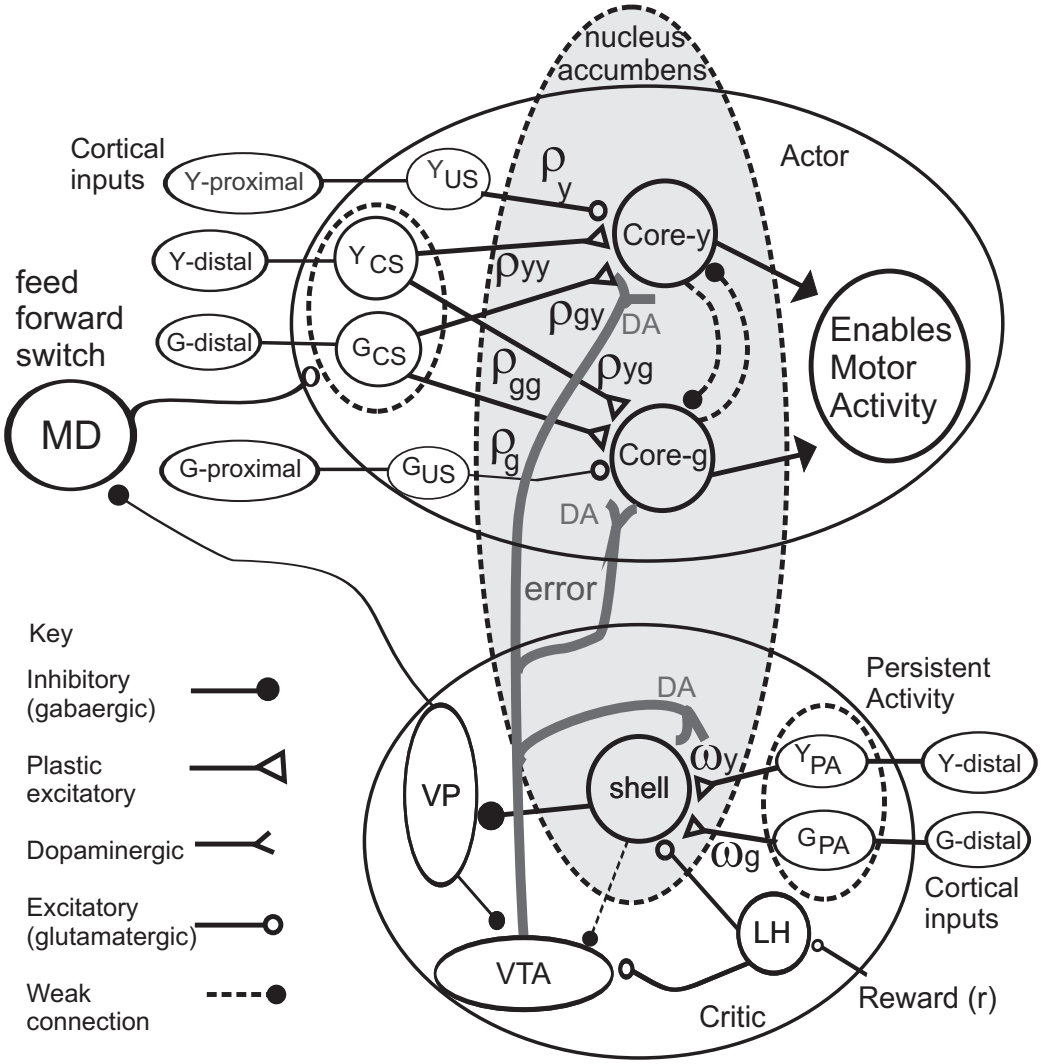

Figure 5.1: The full limbic circuitry model shown as an actor-critic model. Distal and proximal signals from the yellow (Y) and green (G) landmarks represent sensor inputs feeding into the respective cortical inputs that innervate the NAc core and shell unit. The reward activates the lateral hypothalamus (LH) which projects to both the ventral tegmental area (VTA) and the shell. The shell innervates the ventral pallidum (VP) and the ventral tegmental area. The VP innervates the mediodorsal nucleus of the thalamus (MD). The core units use cortical activities to mediate motor behaviours. These cortical afferents to the core are indirectly influenced by the shell via the VP-MD-PFC pathway.

model is illustrated in Fig. 5.2A. The fLTD model functions similarly to the standard actor-critic methods (Sutton and Barto, 1981, 1990) whereby the critic enables the actor to unlearn learned but currently (temporarily) irrelevant associations.

The partial-LTD (pLTD) model

The partial-LTD (pLTD) model illustrated in Fig. 5.2B, is an extended version of the fLTD model. It implements weak LTD on the weights of the actor circuit. Acquired stimulus-response associations are not quickly removed in this model. This model's characteristics suggest that the it should be able to demonstrate rapid reacquisition and rapid responding of acquired behaviours during reversal learning. However, irrelevant responses might persevere because learned associations are not completely disabled when they no longer predict rewards.

The partial-LTD-MD-Feedforward (pLTD-MD) model

The model version that comprises of the MD feed-forward switch and implements a weak LTD rate in the actor so that learned stimulus-action associations are not immediately destroyed is identified as the partial-LTD-MD-feed-forward (pLTD-MD) model and it is illustrated in Fig. 5.2C. The model implemented so far in this thesis represents the pLTD-MD model because it implements weak LTD in the actor circuit and uses the MD feed-forward loop to invigorate and weaken the actor's activity when rewards are presented and omitted respectively.

The distinction between the three models is summarised in table 5.1. In the following sections, these models are compared in both open- and closed-loop behavioural experiments.

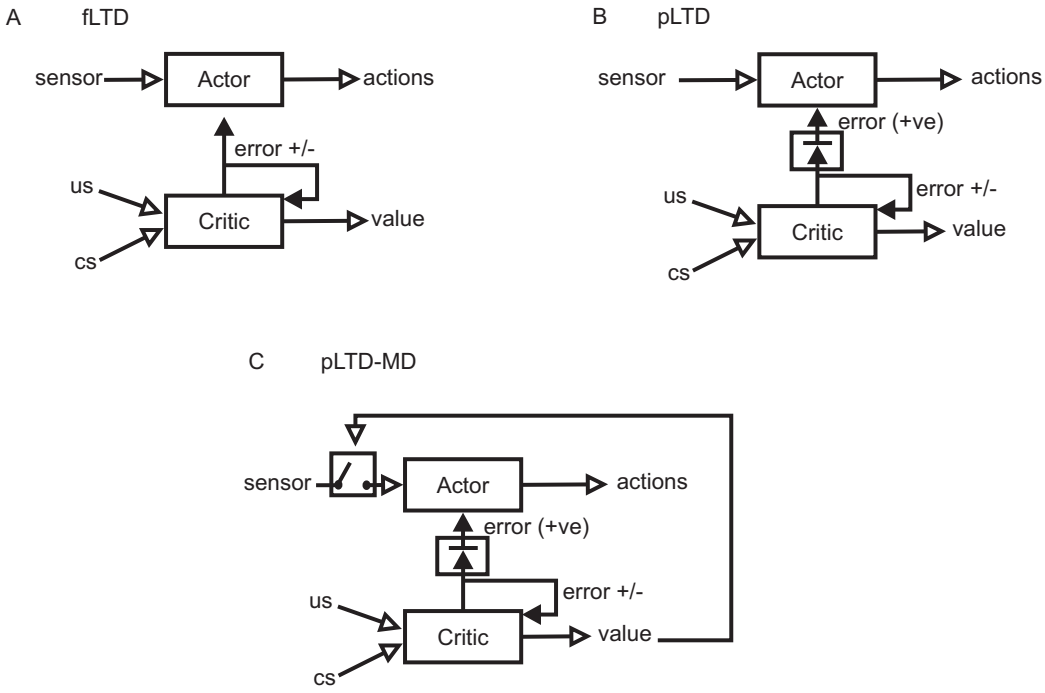

Figure 5.2: A) The full-LTD (fLTD) model B) The partial-LTD (pLTD) model and C) The partial-LTD-MD-feed-forward (pLTD-MD) model

5.2 The Comparison Experiments

The computational model presented in this thesis has two main characteristics which support the idea that stimulus-response (S-R) associations that have been learned, are not immediately eliminated when contingencies change. The first of these two characteristics is that the connectivity between the shell and the core through the shell - ventral pallidal - mediodorsal - prefrontal cortex - core pathway functions as a value switch, which invigorates and disables stimulus-response associations when they result and do not result in reward delivery respectively. The second feature is that there is weak LTD occurring in the core, which is a site in which these stimulus-response associations are formed. The model which implements these two characteristics are well suited to account for rapid reacquisition and perform quite effectively in behavioural serial reversal learning experiments. The following

Table 5.1: The difference between the actor-critic model versions

Model	LTD in the Core (Actor)	Feed Forward MD Switch
fLTD	strong	absent
pLTD	weak	absent
pLTD-MD	weak	present

section illustrates how all three models demonstrate the reacquisition effect.

5.2.1 The Models in Rapid Reacquisition

The reacquisition effect has been tested by subjecting all three models to acquisition and extinction twice as described in chapter 3 over a duration of 100,000 time steps. In the open-loop experiment, delta pulses representing the CS and US with interstimulus intervals of 20 time steps are fed into the model. The magnitude of the outputs of core units (CR) are observed. The response rates of the models during the acquisition and reacquisition stages are illustrated in Fig. 5.3. Here, it can be seen that the CR magnitude of the fLTD and pLTD models are much smaller than the pLTD-MD model which implements the MD feed-forward value switch. The MD invigorates the CR magnitude of the CR responses.

In order to observe the reacquisition effect further, the magnitude of the CR during the initial acquisition and the initial reacquisition when the model is presented with the US after the CS for the first time initially and after the first extinction are shown in Fig. 5.4.

The model can be considered to demonstrate rapid reacquisition. The initial CR magnitude at the beginning of the reacquisition is greater than the initial CR magnitude during the first acquisition. Both pLTD models illustrate rapid reacquisition effects because the S-R associations are not quickly un-

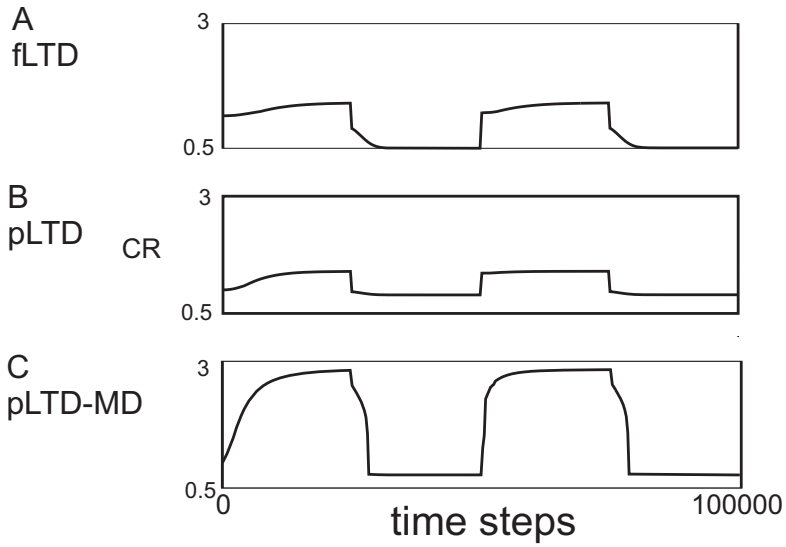

Figure 5.3: The reacquisition effect in the A) fLTD model, B) pLTD model and C) the pLTD-MD model. Simulation parameters are presented in appendix C in table C.1.

learned. In addition the pLTD-MD model unlike the pLTD model, also shows an S-shaped acquisition curve in Fig. 5.3 during both the acquisition and the reacquisition effect. The ability of the pLTD-MD model to demonstrate this trait is favourable in animal learning models (Balkenius and Morén, 1998). The pLTD-MD model shows a greater CR magnitude than the other two models. This indicates how the MD mediates behavioural flexibility by invigorating the action subsystem. The models' performances are also observed in closed-loop behavioural serial reversal learning experiments.

5.2.2 The Models in Serial Reversal Learning

The models were tested in the serial reversal learning experiments as described in chapter 4. In the reversal learning experiments, an agent was required to discriminate between a coloured landmark that contained a reward and one that did not. After this discrimination had been acquired, the contingency was reversed and the agent had to learn to inhibit its original

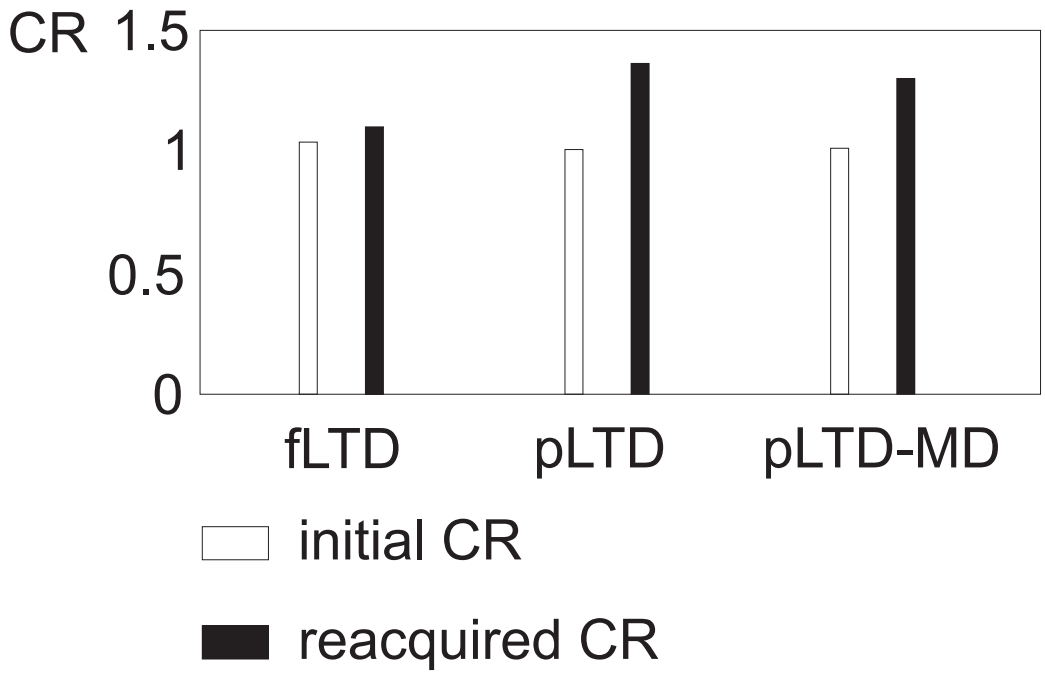

Figure 5.4: A) The initial and reacquired CR in the A) fLTD model, B) pLTD model and C) the pLTD-MD model. Simulation parameters are presented in appendix C in table C.1.

behaviour and learn the new association. This was repeated a few times.

The three models were compared by observing the average speed and average total number of errors the agent made over 10 simulation runs. Fig. 5.5 illustrates the mean total number of reversals the agent achieves over a fixed time duration of 500,000 time steps. It can be seen in Fig 5.5 that the pLTD-MD model achieves more reversals over the fixed time duration than the other two models that do not implement the value switch. The mean errors to criterion produced by each model over an initial discrimination and three reversals are presented in Fig 5.6. It can be seen that the model that utilises the feed-forward switch makes the least amount of errors.

These results indicate that the value switch inspired from biology provides a mechanism by which switching behaviour can more accurately be achieved. The next chapter discusses a variety of actor-critic and non-actor-critic mod-

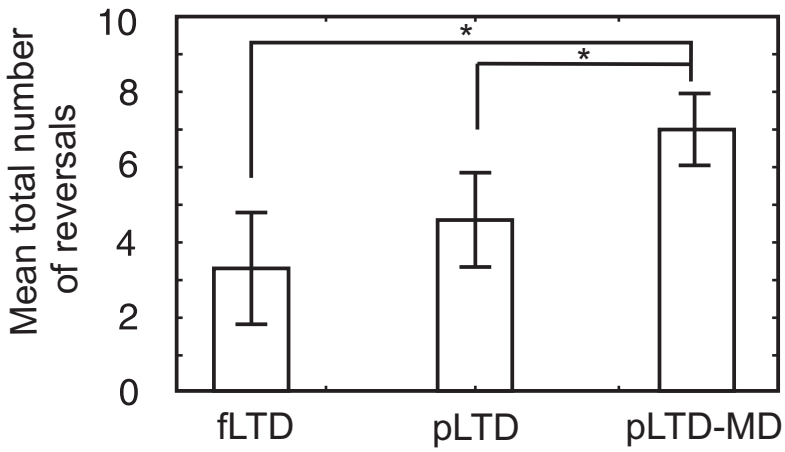

Figure 5.5: The average total number of reversals obtained over a set duration of 500,000 time steps for A) the fLTD (classical actor-critic) model, B) the pLTD (hybrid) model and C) the pLTD-MD (current) model. Average and standard deviation of 10 runs. $*$ indicate statistically significant results for p values < 0.0001. Simulation parameters are presented in appendix C in table C.3.

els.

5.3 Concluding Remarks

In this chapter, the model was represented as a modified actor-critic model and named the pLTD-MD model. It was compared against three different actor-critic versions. The performance of the three models were compared against each other in the rapid reacquisition and serial reversal learning procedures. The results obtained showed that the implementation of a feed-forward switching mechanism, significantly improved the performance of the pLTD-MD model compared to the other two models in demonstrating behavioural flexibility. A variety of actor-critic and non actor-critic models are discussed and compared against the pLTD-MD model in the following chapter. The biological constraints on the model are also discussed.

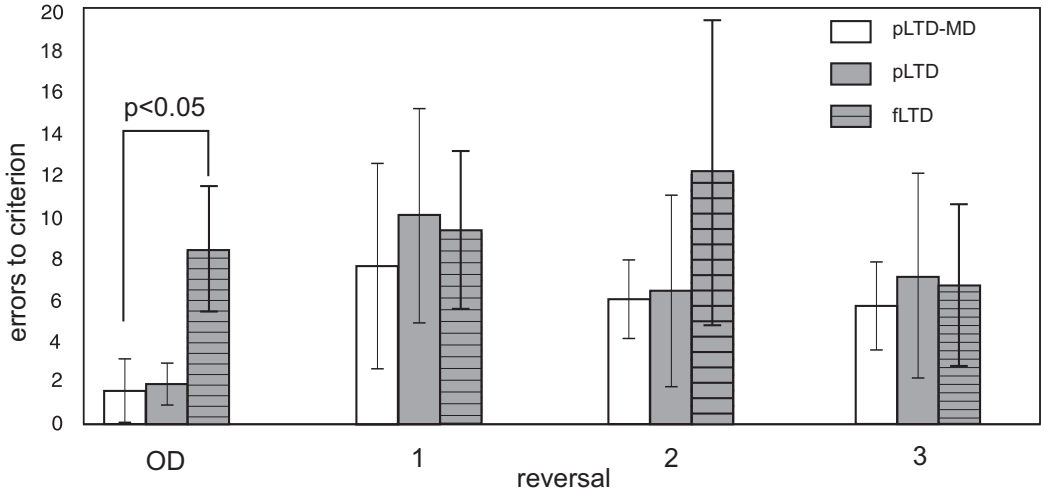

Figure 5.6: The number of errors made by the fLTD (classical actor-critic), pLTD (hybrid) and pLTD-MD (current) models before reaching each criterion to reversal is illustrated for the original discrimination and three consecutive reversals numbered accordingly. Average standard deviation of 10 runs. Statistically significant result has been indicated by the p value. Simulation parameters are presented in appendix C in table C.3.

Chapter 6

Discussion

A variety of neuronal computational models are addressed and compared against the pLTD-MD model presented in chapter 5. The biological constraints are also discussed. The chapter concludes with a summary of the main findings obtained in this work and suggestions for future work. The following section presents a comparative study of the computational model.

6.1 Neuronal Computational Models

A variety of computational models are summarised and compared against the pLTD-MD model in this section. Three issues will be addressed while presenting each model. The first describes what pathways enable DA burst signalling during the CS and US events. The second addresses what processes are involved in reducing the US burst. And the third observes how biologically relevant the first two mechanisms are.

6.1.1 Actor-Critic Architectures

All the actor-critic architectures discussed here use TD learning rule (Sutton and Barto, 1998) which is given by:

$$\delta(t) = r(t) + y(t) - y(t-1) \tag{6.1}$$

This rule has been introduced in chapter 1. The DA signal is calculated by this TD-error so that the DA burst at the primary reward is represented by $r(t)$, and the derivative $y(t) - y(t-1)$ produces the CS burst and the reduction of the US burst during receipt of the expected primary rewards. Each model uses different pathways to the DA producing neurons to explain how the TD-error is calculated. Houk et al. (1995) were among the first to map the actor-critic architecture to the basal ganglia.

Houk et al. (1995)'s Neuronal Critic Model

Houk et al. (1995) present a neuronal model that maps the structure of the cortex and basal ganglia to the adaptive critic. These pathways include the striatum's spiny projection neurons that are classified into two groups, striosomes and matrix modules. The two different striatal modules adopt different roles depending on their characteristic afferent projections. While the striosomal modules which correspond to the striatal striosomes, subthalamic nucleus and DA neurons of the SNc, function as the adaptive critic, the matrix modules assume the role of the actor.

Three main input sources to the SNc which include a direct and indirect pathway are used to generate the firing patterns of DA neurons. The error signal of Eq. 6.1 is produced via the excitatory connectivity from the lateral hypothalamus (LH), the direct projection from the striatal striosome to the SNc and the indirect projection form the striatal striosomes to the SNc via the STN. The direct excitatory LH and slow inhibitory striatal striosome produced the $r(t)$ and $y(t-1)$ components respectively while the indirect

striatal striosomal - subthalamic nucleus pathway which resulted in a net excitatory effect was mapped as the $y(t)$ component.

Houk et al. (1995)'s use of persistent slow inhibition of DA cells via the direct pathway is not consistent with data because this process predicts a sustained depression in dopamine cell firing from the time of the CS presentation to the time the reward is obtained. According to data, the DA bursts are generated at the time of the CS onset after which the DA activity settles to baseline and does not show persistent inhibition. In addition during omission, the DA neuron activity only deviates for a short duration from the baseline (Brown et al., 1999). By projecting persistent inhibition, the model does not account for the exact timing of the observed depression when the reward is omitted (Joel et al., 2002). A second inconsistency is observed in the direct projection of the striatal-SNc projection. The efferents of the striatum are inhibitory therefore the indirect excitatory pathway to the SNc via the STN would result in inhibition rather than excitation so that the $y(t)$ component becomes $-y(t)$ (Porr and Wörgötter, 2005).

The model implemented by Suri and Schultz (1999) is extended from the actor-critic model of Barto (1995) which addresses the temporal aspects involved in predicting expected rewards.

Suri and Schultz (1998, 1999)'s Model

Suri and Schultz (1998, 1999) present a neural network actor-critic model that was trained to perform a simulated spatial delayed response task. Rather than implement a delayed prolong inhibition, Suri and Schultz (1998, 1999) reproduce the timing mechanisms involved in the depression of DA activity during the event of the omitted reward by representing a stimulus as a series of signals activated over different durations.

While a substantial amount of effort was made in which a modified version of the critic was used to replicate the firing patterns of DA neurons, little attempt was made to map the modified rule to the basal ganglia architec-

ture. However, Suri and Schultz (1998, 1999) briefly suggest that reward predictions that are learned are mediated by the projections of the cortex to the DA neurons via the striatal patch striosome compartments. They also proposed that DA neurons or the striatum are the potential sites in which the time derivative of the prediction may be computed. The TD algorithm was extended to account for novelty and generalization responses as well as temporal aspects by adjusting and setting the parameter values without providing a biological assumption or justification. Contreras-Vidal and Schultz (1999) make an attempt to provide a model based on the basal ganglia architecture that accounts for DA responses to novelty, generalisation and appetitive and aversive stimuli.

The computational models discussed so far have focussed mainly on the basal ganglia nuclei. In particular, these models do not specifically aim to reproduce a model that focusses on the ventral striatal circuitry and which describe the role of the core and shell sub divisions in motivation and reward related learning. In an attempt to address motivational processes, the limbic system has been modelled by Dayan (2001).

Dayan (2001)'s Model

Dayan (2001) presents an extended actor-critic model that accounts for the motivational processes which dictate whether a single action is worth executing. The model implements the prediction error represented by Eq. 6.1.

In the model, $r(t)$ is determined by the "*hard-wired*" US evaluator. The US signals could also be calculated via a plastic route. $y(t)$ was calculated from two competing sources. The first included both the basolateral nuclei of the amygdala (BLA), and the orbitofrontal cortex (OFC) which contained a prior bias. The second competing source was largely dependent on a stimulus substitution relationship between the CS and the US.

The shell accounts for Pavlovian motivation for pre-wired and new habits. It is trained by the error signal and it determines the vigor by which an action

is executed. Its activity also influences the $y(t)$ term. An advantage function is implemented with the core to control the instrumental motivation for the action choice. The advantage is also trained by the TD error.

All of the above models have linked the TD learning algorithm to DA activity. However, these TD methods have not fully or accurately specified what biological processes actually produce fast excitatory responding to the CS as well as a delayed adaptively timed inhibition of response towards the US. Brown et al. (1999) use known anatomy and pathways to present a model which aims to explain how the DA signal is generated.

6.1.2 Other Computational Models

The models discussed so far fall under the category of actor-critic models of reinforcement learning. However, there are numerous computational models which do not utilise the classical actor-critic architecture one of which is that implemented by Brown et al. (1999).

Brown et al. (1999)'s Model

Brown et al. (1999) present a model that proposes two parallel direct and indirect pathways from the cortex to the DA system of the SNc. These pathways perform either excitatory or inhibitory conditioning which generate the DA response at the CS and inhibit the DA response during the US respectively. In the model, the excitatory pathway involves the pendunculopontine tegmental nucleus (PPTN) connection to the DA neurons of the SNc. The lateral hypothalamus (LH) which becomes activated when primary rewards are obtained activate the PPTN which in turn project to the SNc. This pathway is responsible for the DA burst in event of primary rewards. On the other hand, the CS induced DA burst involves the pathway from the limbic cortex which excites the ventral striatum which inhibits the ventral pallidum that produces inhibitory projections to the PPTN.

The adaptively timed inhibitory conditioning, implements a mechanism which involves glutamate receptor driven Ca^{2+} spikes generated within a spectrum of temporal delays. Inhibitory learning is enhanced when the Ca^{2+} spike is activated at the same time as a DA burst. This inhibitory conditioning implements the $Ca2+$ spectral timing mechanism which are activated by the CS projected by the limbic cortex to the striosomes. The CS traces project via adaptive pathways to both the ventral striatum and the striosomes. Learning occurs in both the excitatory and inhibitory pathway. The CS representation learns to drive the DA burst and activate the $Ca2+$ spikes in the striosomal cells which learn to inhibit the DA bursts during the primary reward and cause a dip when rewards are absent (Brown et al., 1999).

Another model that implements learning at pathways other than the corticostriatal synapses is implemented by Berns and Sejnowski (1998).

Berns and Sejnowski (1998)

Berns and Sejnowski (1998) presented a systems-level model function of the basal ganglia that accurately mapped the connectivity of the anatomical structures. Unlike previous models, the model demonstrated an ability to learn to reproduce action sequences.

Actions were selected by a release of tonic inhibition of the thalamus via an inhibition of the globus pallidus (GP) which inhibits the thalamic nucleus. This was achieved by the (direct) inhibitory connection from the striatum on the GP which resulted in a "*loser take all*" mechanism of action selection.

In order to demonstrate sequential action selection effectively, an ability to store memory was required. The model achieved this by assuming that the feed-back loop between the subthalamic nucleus (STN) and the external segment of the globus pallidus (GPe) generated a form of short term memory. Unlike other computational models, learning in the form of 3 factor Hebbian learning occurred in the synapses between the striatum and the GP and the STN and the GP. The GP represented the postsynaptic component while

the respective presynaptic components were the striatum and the STN. The third factor was the error signal originating from dopamine neurons which calculated the differences between the striatum-GP and the STN-GP activations.

The dis-inhibition of thalamic inputs as well as the use of 3-factor Hebbian learning, are ideas that have also been implemented in the pLTD-MD model. However, learning occurs on cortico-striatal synapses rather than on the afferent synapses to the GP. Also, while the pLTD-MD model has been tested in reversal learning experiments rather than in performing sequential learning, the pLTD-MD model is also capable of performing in sequential learning tasks. The model can acquire associations between secondary CS and primary CS by employing DA activity that develops earlier in event of the CS (Fig. 3.10) (Thompson et al., 2008).

Schmajuk et al. (2000)'s model focusses on the NAc circuitry and uses it to describe latent inhibition.

Schmajuk et al. (2000)'s Model

Schmajuk et al. (2000) propose a mechanism that accounts for latent inhibition (LI) using a modified version of the Schmajuk, Lam and Gray (SLG) model (Schmajuk et al., 1996) for classical conditioning. Latent inhibition (LI) is generated when the learning of conditioned associations to a stimulus is retarded due to prior exposure to the stimulus. The model is mapped onto the nucleus accumbens (NAc), central nucleus of the amygdala (CNA), hippocampus, enthorhinal cortex (EC) and DA neurons and used to demonstrate the impairment, restoration and preservation of LI by simulating lesion effects (Schmajuk et al., 2000; Schmajuk, 2005).

The model employs a novelty feature which calculates a mismatch between the predicted and observed events of the CSs, contexts and the US. Novelty is represented by dopamine (DA) in the NAc. The EC computes predictions of CSs and USs and the contexts and projects the information directly

to the hippocampus, shell and indirectly to the VTA via the shell. There are two components of novelty which are computed by the VTA. The first component is calculated form the excitatory information it receives from the pendunculopontine tegmental nucleus (PPTN) about the observed values and information from the NAc shell about predictions. This component makes calculations when the average observed values exceed the average predicted values. The second component of novelty is computed when the average predicted values exceed the average actual values. It is calculated by the inhibitory information from GABAergic afferents which include signals from the NAc. The model proposes that the novelty signal computed in the VTA projects to the NAc core which in turn relays a signal proportional to novelty to the thalamus via the VP. The thalamus is associated with the internal representation of the CS. This internal CS representation is used to form a CS-US association in the CNA and reflects the strength of fear conditioning (Schmajuk et al., 2000).

The model is used to describe LI as follows: When the model is pre-exposed to the CS, predictions of the CS increase resulting in increased activity in the shell, decreased activity in the VTA, the first component of novelty, and the thalamus. This results in a retarded formation of CS-US associations in the CNA. Schmajuk et al. (2000) simulate adjustments in the model so as to represent lesioning effects and the application of DA agonists and antagonists. They show how the model describes impaired LI by shell lesions, restoration of LI by the application of haloperidol, preservation of LI by core lesions, facilitation of LI by combined shell and core lesions. Impairments of LI due to hippocampal or EC lesions are in turn restored by haloperidol.

Another model that does not utilise the TD algorithm but implements the Rescorla-Wagner rule instead is the primary value learned value model developed by O'Reilly et al. (2007). Like Dayan (2001)'s model, O'Reilly et al. (2007) propose a model which focuses on the ventral striatum.

O'Reilly et al. (2007)'s Model

The primary value learned value (PVLV) model (O'Reilly et al., 2007) although not as computationally elegant as the TD model, was developed so as to produce a more relevant biophysical model for Pavlovian learning that also accounts for and maps more accurately to the reward-predictive behaviour of midbrain dopaminergic neurons and performs more robustly to a variable environment than the TD model (O'Reilly et al., 2007).

The PVLV model comprises two subsystems namely the primary value (PV) and the learned value (LV) systems both of which utilise the Rescorla-Wagner or delta-rule. The PV system is used to learn the occurrence of primary rewards while the LV system is used to train the secondary CS-CS association.

The PV system is activated by primary rewards which correspond to the excitatory LH projection to DA neurons. This excitatory activation is inhibited by the NAc which is used to suppress DA bursts from the primary reward. This connectivity is in accordance with the pLTD-MD model. However, unlike the pLTD-MD model which uses the indirect NAc-VP-VTA pathway to account for activation during the CS onset, the LV system is activated by the central nucleus of the amygdala (CNA) which activates the DA burst during the CS onset. Like the pLTD-MD model and PV system, the LV receives inhibitory GABAergic influence from the NAc which slowly removes the DA burst that occurs during the CS onset. In the pLTD-MD model, the NAc activity on the VTA via direct and indirect pathways are sufficient to respectively suppress DA bursts during the US and enhance the DA burst during the CS. The LV learns on the condition that primary rewards are expected or available. In the pLTD-MD model, this condition is not necessary because of the direct LH-NAc connectivity which bootstraps learning. Learning occurs in the NAc which in turn plays a role in DA activity. Therefore the NAc influences DA release and as with both the LH and feed-back mechanism, its own "*self learning*". Unlike the condition set on the LV system, no external conditions are required to determine how learning occurs in the pLTD-MD model.

The next section presents a general discussion in which some of the models presented above are compared against the pLTD-MD model.

6.1.3 Discussing the Models

The actor-critic models described so far implement the TD algorithm which uses one clean equation to predict both current and future rewards at the US and CS events respectively. This means that the models are capable of computing associations between primary CS-US links and higher order CS-CS associations. However, a serial unbroken chain or a precisely timed representation between both higher order secondary and primary stimulus is essential for the reward prediction error (DA burst) to gradually propagate to the earliest occurring CS. The PVLV model does not depend on a linked serial compound representation of all the stimuli so as to acquire such associations O'Reilly et al. (2007). The learned value's dependence on the primary value means that the model is limited to second order conditioning only. This is not the case in classical TD methods and the model presented here.

In most of the models discussed, the error is calculated in the DA neurons and delivered globally so that weights increase or decrease in an identical manner depending on its value. In the pLTD-MD model, there are two DA transmission modes which are also released globally but influence weight change on the target structures uniquely depending on the targets surrounding synaptic activities (Malenka and Bear, 2004). The two DA transmission modes are produced in the pLTD-MD model as follows: A reward delivery generates DA bursts which produce phasic levels of dopamine. An omission of expected rewards results in tonic DA levels which are generated when the shell activity dis-inhibits the VTA through the VP. Although also released globally as well as on the NAc, the phasic DA activity is used to signal when, rather than how much learning should occur. The weight increase itself is dependent on pre and post synaptic activity in NAc. This is extremely useful for localising learning because DA neurons project to a variety of brain regions. This means that tonic and burst DA activity can be used to encode

different effects in different brain regions. Schmajuk et al. (2000) suggest that rather than coding for reward, the DA neurons code for the magnitude of novelty.

The O'Reilly et al. (2007), Dayan (2001) and the pLTD-MD model propose that learning occurs in the NAc which has a direct contribution to the response. In contrast, Schmajuk et al. (2000)'s model proposes that the NAc simply contributes to novelty calculated in the VTA which indirectly contributes as the internal representation of the CS. The CR is generated when an association between an internal representations of the CS and the US is obtained in the amygdala.

An elimination of learned associations during extinction in current computational models (O'Reilly et al., 2007; Dayan, 2001) implies that a similar rate to relearn the association is required when the US is reintroduced. This process does not account for the rapid reacquisition which have been observed to occur more quickly than the original acquisition (Pavlov, 1927; Napier et al., 1992). The model presented here inhibits rather than removes unnecessary learned behaviour so that when contingencies change and once the previously irrelevant behaviour becomes useful again, it is no longer suppressed and can very quickly be reinstated.

LTD encoded in the pLTD-MD model in event of an increased DA activity, occurring due to a rise in the number of tonically active neurons. In the PVLV and TD methods, a negative prediction error is encoded by a pause in the tonically firing DA neurons. The problem with using a negative value to calculate the error signal during omission is that the negative error does not quantitatively correlate with the pause in DA activity (Cragg, 2006). In addition, due to the low baseline firing rates characterised by DA neurons, it might be difficult for the recipient structures to decode a pause in tonic firing (Cragg, 2006; Daw et al., 2002). By employing two different levels of increased DA transmission to encode both LTP and LTD, the problem of generating a negative error value dependent on this pause in DA activity is avoided. The method by which DA activity is encoded here ensures that the

necessary process required for weight change is distinctly identified. Although the shell has both an inhibitory and dis-inhibitory effect on the VTA via a direct and indirect pathway, the direct pathway seems to have a weaker effect than the shell-VP-VTA pathway (Zahm, 2000). Thus an increase in the tonically active DA neurons would seem to occur more readily than a pause in activity.

6.2 Discussing the Biological Constraints

The current model proposes that the dopaminergic neurons of the VTA are activated via a direct excitatory glutamatergic pathway, and an indirect dis-inhibitory GABAergic pathway. These pathways generate a burst in DA neurons and an increase in the population firing of tonically active neurons. The burst and tonic DA activity respectively mediate weight increase or long term potentiation (LTP) and weight decrease or long term depression (LTD). Acquisitions occur when unexpected rewards are obtained. When rewards are omitted, rather than implement a pause in tonic firing patterns (Dayan and Balleine, 2002; O'Reilly et al., 2007), the current model employs a rise in tonic activity to to encode reward omission. In the current model, DA activity mediates weight change however, there are many roles which DA activity has been implicated in some of which are discussed next.

6.3 The Role of Dopamine Activity

Dopamine plays a central role in learning (Robbins and Everitt, 1996). It is essential for mediating plasticity in the NAc has been implemented as an error signal in numerous computational models (Suri and Schultz, 1999; Houk et al., 1995; Joel et al., 2002). It has also been proposed to facilitate and attenuate synaptic transmission in the dorsal striatum (Prescott et al., 2006).

The dorsal striatum is divided into two different populations (Gerfen, 1988), depending on their ability to express dopamine receptors. These two populations differentially express D1 and or D2 receptors and constitute the direct and indirect pathways respectively (Clark and Boutros, 1999; Parent et al., 2000). With the DA receptors responding distinctly to different DA transmission levels, the hypothesis is that the dopaminergic activation of D1 receptors results in enhanced synaptic efficacy while the opposite is observed when DA activates D2 receptors (Clark and Boutros, 1999). These mechanisms have been used to suggest how DA subserves switching behaviour Redgrave et al. (1999a); Prescott et al. (2006); Gurney et al. (2001a,b). The most popular interpretation of transient DAergic activity, is that transient DAergic activity signals the error in the prediction of future rewards. Redgrave et al. (1999b); Redgrave and Gurney (2006) have suggested that this short-latency DAergic activity plays a role in the switching of attentional and behavioural selections to unexpected relevant stimuli. In the current model, transient DA activity has been used to indicate the relevant moment *when* learning should occur. Although Redgrave et al. (1999a) suggested that rather than learning, DA played a role in switching behaviour, Weiner and Joel (2002) supported the view that DA is involved in both learning and switching behaviour by strengthening corticostriatal synaptic plasticity and attenuating and facilitating corticostriatal transmission respectively (Joel et al., 2002). The current model proposes that while DA mediates learning in the ventral striatum, the shell-MD-VP-core pathway is required for behavioural flexibility and therefore, plays a role in switching behaviour.

There are a variety of roles which tonic DA activity are suggested to be involved in. For instance, due to the elevated DA levels which have been observed to occur in response to aversive stimuli (Horvitz, 2000; Salamone et al., 1997), Daw et al. (2002) proposed that tonic DA levels signal average punishment. On the other hand, based on the link between tonic DA levels and energized behaviour, Niv et al. (2007) suggested that this DA activity encodes the average reward rate signal useful in exerting control over the vigor of responses. By manipulating different regions of the accumbens, Reynolds

and Berridge (2001) observed both positive and negative motivational behaviours. The variety of functions tonic DA has been associated with along with the diverse behaviours the NAc seems to be involved in mediating suggests that DA release on this structure could occur at different rates (Barrot et al., 2000; McKittrick and Abercrombie, 2007) or could produce varied effects dependent on the target discharge sites.

It has been suggested that phasic and tonic DA respectively mediate LTP and LTD according to the following findings: Phasic and tonic DA activity produces different DA concentration levels. According to Pawlak and Kerr (2008), the function DA receptors play on synaptic transmission is dependent on this DA concentration. DA bursts generate higher levels of DA which activate D1 receptors and induce LTP. On the other hand tonic DA levels stimulate D2 receptors which play a role in mediating LTD (Calabresi et al., 1992b). Additionally, tonic DA exerts different effects on the shell and the core such that LTD occurring in the shell is significantly stronger that LTD in the core. These assumptions need to be validated empirically. This can be done by observing synaptic plasticity when these specific regions are manipulated by either DA D1 and D2 receptor agonists and antagonists, or by DA applications and depletions. Studies conducted by (Churchill et al., 1996) have demonstrated that the application of GABA agonists on the mediodorsal nucleus of the thalamus resulted in increased locomotion. This behavior was related to a decrease of DA activity in the PFC. In addition, an increase in DA activity was observed in the core region but not the shell or dorsolateral striatum (Churchill et al., 1996). These studies indicate how different DA activities and therefore distinct effects are generated in the NAc sub-regions.

LTP occurs in the model NAc core and shell through three factor Isotropic sequence order learning (ISO-3) (Porr and Wörgötter, 2003). The third factor corresponds to DA burst which gates synaptic plasticity. During omission, the absence the DA bursts along with extended tonic activity due to the prolonged CS influences results in stronger LTD in the shell. LTD is produced in the shell when there is pre-synaptic activity occurring in concert with DA tonic activity. Studies from Pawlak and Kerr (2008) have shown that D1

but not D2 receptor activation is necessary for STDP. Although the current work requires D1/D2 receptor activation to induce LTP/LTD respectively, D1 receptors are also capable of enabling LTD. The model utilises a form of ISO learning which has been shown to generate LTP and LTD depending on the timing between the pre- and post-synaptic activities. If the pre-synaptic activity occurs after post-synaptic operation, LTD can be induced. The 3rd factor (D1 receptor activation) simply enables such spike timing dependent plasticity (STDP). This means that the D1 receptor is sufficient to enable LTD through STDP. However in addition to this, D2 receptor stimulation dependent on tonic DA concentration levels is also capable of inducing LTD.

According to Fiorillo et al. (2003), tonic DA levels seem to carry information about the uncertainty of rewards whereby they exhibit highest levels when rewards are delivered with a probability of 0.5 and lower levels at probabilities tending towards 1 or 0. This might indicate that these varying DA levels which seem to encode further information about rewards differentially influence synaptic transmission. This work does not account for intermediate levels of DA and the possibility that LTP and LTD induction might in addition, be sensitive to these different intermediate DA concentration levels (Matsuda et al., 2006), although it suggests that such specific DA concentrations might provide favourable conditions that prepare the synapse for both LTP and LTD so that any one can very quickly be induced. This DA level could be associated with the observed sustained activation of DA neurons that precede uncertain rewards (Fiorillo et al., 2003) so that when reward delivery or omission becomes more certain, the levels readjust accordingly.

DA bursts are generated in event of the CS and US bursts. While the LH activates bursts in receipts of rewards, The CS bursts are produced via VTA dis-inhibition by the shell through the VP. The direct shell-VTA pathway has been modelled to have a stronger inhibitory influence on the VTA than through the VP only when the weights in the shell have developed to a level such that the shell activity becomes stronger that the VP inhibition. As the shell weights develop, the bursts occurring in event of the US start to decrease, eventually DA bursts which switch on learning when rewards are

obtained are no longer generated. However, bursts occurring at the onset of the CS generated through the dis-inhibition are useful for both secondary conditioning (Thompson et al., 2008) and reducing the early induction of LTD. The modifications implemented in the current model should not limit its ability to perform secondary conditioning, but provide added versatility by enabling behavioural flexibility through the suppression of unnecessary acquired actions.

6.4 The Role of the NAc

Central to the model is the NAc shell and core. These two sub units function in distinct manners whereby the acquisition for stimuli that precede reward delivery occurring in the core, enables instrumental responding to reward predictive cues. On the other hand, both acquisition and depression occur in the shell when rewards are obtained and are omitted respectively. This allows for a mechanism by which the shell activity is quickly updated depending on the presence or absence of rewards and enables switching behaviour by indirectly suppressing activity in the core. Activity in the shell results in the dis-inhibition of the MD and VTA via the VP with the former altering instrumental responding controlled by the core. Another function of the shell is to influence the population of tonically active DA neurons in the VTA. This occurs through dis-inhibition of the DA neurons via the VP (Floresco et al., 2003).

Although the shell as a value system has been accepted and implemented in theoretical models, a novel biological functionality of the shell has been added such that the shell is also capable of facilitating and attenuating the input system to the actor (core) so that learned associations in the actor are not eliminated. The value of reward predicting stimuli are updated in the shell which inhibits the behaviour towards the previously relevant stimulus through the Shell - VP - MD - PFC -core loop (Zahm and Brog, 1992; Birrell and Brown, 2000). LTD in the shell results in reduced shell activity and

increased inhibition of the MD via the VP. This produces an attenuated cortical activity to the core and a resultant attenuation of behaviour. Thus learned behaviour towards the now irrelevant stimulus is inhibited or gated. If the stimulus - reward contingency switches again, the inhibited behaviour is quickly dissolved as LTP is quickly reinstated in the shell again and the MD is dis-inhibited. Therefore, the shell (value system) modulates the PFC which processes the stimuli that predicts the availability of a reward.

There are a number of biological experiments which substantiate the model presented here, a few of which are discussed as follows: The shell and core have been identified to play distinct roles when responding to reward predictive cues. Accordingly, lesion experiments conducted by Floresco et al. (2008a) suggest that the shell facilitates alterations in behaviour in response to changes in the incentive value of the conditioned stimuli, while the core allows reward predictive stimuli to enable instrumental responding. The flexibility demonstrated by the shell with respect to the changing value of incentive value could occur due to LTP and LTD occurring mainly in the shell.

The current work suggests that LTP and LTD are influenced through the activation of D1 and D2 receptors respectively. According to Calaminus and Hauber (2007), DA transmission on the NAc which activates D1-like and D2-like receptors is essential for generating response to reward predicting cues. Also, Cools et al. (2007) have observed that dopaminergic modulation in the nucleus accumbens plays a role in reversal learning. However, experiments done by Calaminus and Hauber (2007) suggest that D1 and D2 receptor activation on the core while mediating instrumental behaviour, is not crucial for updating the incentive values of reward predictive cues. A blockade of DA receptors on the OFC has been observed to impair reversal learning (Calaminus and Hauber, 2008). These findings support the proposed model in which it is suggested that D2 receptor activation plays an important role in enabling LTD on the OFC afferents to the shell. Accordingly the shell seems to be the more relevant nucleus required in updating the incentive values of conditioned reinforcers. On the other hand, more recent findings

have shown that D2 receptor agonists applied to the core in a dose dependent manner impaired reversal learning by significantly increasing the preservative errors. This increase in preservative error may occur because the elevated D2 agonist generates stronger resultant LTD than LTP in the core, such that new associations can not be learned and original learned actions persevere.

Lesion and inactivation studies on the shell as opposed to core inactivation results, have shown that the shell seems to have an inhibitory effect on behaviour (Blaiss and Janak, 2008). While there is very little evidence which show strong direct connectivity between the shell and the core, the inhibitory effect of the shell on behaviour can be explained by the indirect activation of the cortical afferents to the core via the MD. This pathway allows the strong cortico-striatal activation of one specific core neuron to inhibit other competing core neurons. The distinct roles of the NAc shell and core subunits have been documented and implemented in a computational model which has successfully demonstrated behavioural flexibility in a reversal learning food seeking procedure.

The prelimbic area in the rat prefrontal cortex which innervates the core (Brog et al., 1993) plays an essential role in initiating reward or drug seeking behaviours (Ongür and Price, 2000; Peters et al., 2008). Studies conducted by Peters et al. (2008) suggest that shell as well as the infralimbic area of the PFC which projects to the shell (Brog et al., 1993), are recruited by extinction learning to suppress reward seeking behaviour. Lesion and inactivation studies on the shell compared to the core, has shown that the shell seems to have an inhibitory effect on behaviour (Blaiss and Janak, 2008). While there is very little evidence which show strong direct connectivity between the shell and the core, the inhibitory effect of the shell on behaviour can be explained by the indirect activation of the cortical afferents to the core via the MD. This pathway allows the strong cortico-striatal activation of one specific core neuron to inhibit other competing core neurons. The shell influences the MD which in turn produces a reduced activity on the cortical afferents to the core.

Overall, these studies are a few among many which suggest that the NAc functions as an important interface through which the motivational effects of reward predicting cues and stimuli obtained from limbic and cortical regions transfer onto response mechanisms and instrumental behaviours (Di Ciano et al., 2001; Cardinal et al., 2002a,b; Balleine and Killcross, 1994).

6.5 The Role of the MD Thalamus

The shell dis-inhibits the MD through the shell-VP-MD pathway. It has been shown that the activation of the MD or the inhibition of GABA receptors in the MD generates elevated DA activity in the PFC (Jones et al., 1987, 1988). An inverse relationship between dopamine transmission in the PFC and the NAc and has been observed (Deutch, 1992). This means that the dis-inhibition of the MD which results in enhanced DA activity in the PFC generates reduced activity in the NAc core (Churchill et al., 1996). This reduced DA activity in the core might result in a facilitation of PFC inputs due to the resultant reduced DA-D2 receptor activation (Goto and Grace, 2005b). Such facilitation of PFC inputs might enhance the ability of the overall circuit to select active sensor inputs.

The MD plays an additional role by providing the current model with the added characteristics of being robust. While the current model requires that LTD in the core occurs at a significantly lower rate so as to ensure that learned actions are maintained, during omission of rewards, when the activity of the MD is reduced, the PFC inputs to the core are also attenuated. This limits the amount by which LTD is generated in the core. LTD occurs in event of both tonic DA levels and pre-synaptic activity. The MD's indirect influence on the rate of LTD on the core can be observed from the learning rule implemented in the model Eq.4.10 as developed in chapter 4

$$\begin{aligned}\rho_X(t) \leftarrow \rho_X(t) \quad &+ \quad \mu_{core}(X_{CS}(t) \cdot \textit{core-}X(t)' \cdot burst(t) \cdot (1 - \rho_x(t))) \\ &- \quad \epsilon_{core}(X_{CS}(t) \cdot tonic(t))\end{aligned}$$

To recap, the weight (ρ_X) increases (LTP occurs) when there is a correlation between the presynaptic activity ($X_{CS}(t)$), the postsynaptic activity ($\textit{core-}X(t)'$) and a dopamine burst. The negative part of the equation represents weight decrease or LTD in the core and occurs when there is a correlation between tonic activity and the pre-synaptic activity ($X_{CS}(t)$) which in turn is influenced by the MD i.e. Eq.4.11 in chapter 4 is illustrated below:

$$X_{CS}(t) = u_{\textit{X-distal}}(t) + \theta_{MD} MD(t)$$

By shutting down pre-synaptic activity to the core the MD also indirectly reduces the rate of LTD in the core. This suggests that the MD improves the robustness of the model in maintaining established stimulus-response associations. The functional link of the MD thalamus on the PFC in the association of stimulus responses is substantiated by the similarities observed by Chudasama et al. (2001) which demonstrated reversal learning impairments following MD thalamus and mPFC lesions. Errors were observed in MD lesioned agents not during acquisition, but during the reversal of stimulus-reward contingencies. These findings are consistent with results obtained by (Means et al., 1975) who observed increased perseverative errors in reversal learning tasks performed by agents with thalamic lesions. The above studies work in concert with the current model in which during reversal, LTD occurring in the shell influences the responses mediated by the core through reduced inhibition on the MD thalamus.

6.6 The Model in Latent Inhibition

Experimental studies have shown that increased and decreased DAergic activity on the NAc resulted in enhanced switching and perserverative behaviours respectively (Weiner, 2003; Taghzouti et al., 1985b) In particular, potentiated LI was observed in animals with either DA depletion in the NAc or application of D2 receptor antagonists (Joseph et al., 2000). Application of amphetamine which generates DA over-reactivity resulted in disrupted LI (Weiner, 2003). In addition, while LI was left intact or persistent under conditions that disrupt LI, in core lesioned agents; It was disrupted in shell lesioned agents.

Disrupted LI due to amphetamine can be explained in the model as follows: amphetamine generates a resultant increase in DA which favours acquisition in the current model. CS-US associations are obtained more readily so that switching behaviour occurs in response to CS generated during high levels of DA and LI is attenuated. In addition, the DA levels might also be activating D2 receptors which according to the model also enables LTD. LTD and LTP might be occurring simultaneously in the shell so that the resultant increased and decreased activity influences the core and therefore the CR generated through the feed-forward pathway. The model proposes that intact or persistent LI observed due to core lesions can be described as follows: Stimulus-response (S-R) associations are acquired in the core which if lesioned would result in impaired S-R conditioning. This would lead to attenuated responding to stimuli which could reflect the intact/persistent LI. On the other hand shell lesion which results in disrupted LI might occur because the shell is implicated in inhibiting the ability to switch. This might occur through its indirect influence on the core via the thalamus. It is proposed that the shell provides a mechanism by which general responses enabled by the core are made specific through the feed-forward pathway. In addition, the MD-PFC-striatal circuitry has been observed to subserve certain types of working memory (Floresco et al., 1999). This means that lesions to the shell may result in reduced activity in the MD-PFC-striatal pathway and

therefore a depleted ability to retain a short-term memory of information about recent stimuli and its relevance. This effect might in turn favour the switching of behaviour.

6.7 Summary of Main Findings

This thesis began with an introduction to adaptability in control systems and embedded agents. Animal learning was introduced as open- and closed-loop procedures. A few neuronal analogs of adaptation and animal learning were briefly summarised. This began with a neuronal representation for Hebbian learning and ended with the actor-critic algorithms. Actor-critic methods have been used to suggest how the basal ganglia and the cortex perform prediction and control (Houk et al., 1995; Joel et al., 2002). This chapter demonstrated how adaptive behaviour and animal learning can be modelled using different computational methods. Some computational models of the basal ganglia which have been implicated in a range of psychomotor behaviours were introduced briefly. Numerous algorithms including classical actor-critic methods implemented as computational models for animal learning assume methods in which learned associations are destroyed when they no longer predict rewards. These techniques seem to be biologically inefficient processes for demonstrating behavioral flexibility. The sub-cortical limbic system and its dopaminergic innervation as a substrate for reward based learning was briefly introduced.

In chapter 2, the basal ganglia was reintroduced with the striatum described in terms of its dorsal and ventral division. The dorsal striatal region was summarised however, a more elaborate description of the circuitry of the ventral striatum was provided. The study focused on the role the nucleus accumbens (NAc) shell and core played in behavioral flexibility and reward based behaviours. Finally, a few characteristics and assumptions of the ventral striatum were obtained and used to generate a computational model. Suggestions were made regarding the role of the shell and the core. The shell

has been assumed to mediate change in behavior with respect to the incentive values of the conditioned stimuli. The core plays a role in enabling reward predicting stimuli to mediate instrumental responding. Further assumptions were made about the rate of cortico striatal plasticity in the shell and the core. In particular, LTD occurs more significantly in the shell than in the core. The shell is also proposed to influence activity in the core through a shell - ventral pallido - thalamo - cortical - core pathway.

The model developed assumes that plasticity in the shell and core are mediated by dopaminergic activity. Tonic and burst dopaminergic activities respectively act on D2 and D1 type receptors which facilitate LTD and LTP respectively.

A computational model of the NAc circuitry based on the studies discussed in chapter 2 was produced in chapter 3. The model was developed using a form of differential Hebbian learning. In addition, the model was tested as an open-loop system whereby the systems response was not conditional on the outcome. This showed that the model performed in accordance with a range of classical conditioning phenomena.

In chapter 4 the model was tested in a variety of closed-loop behavioural reward seeking experiments. These closed-loop experiments included a simple reward seeking task in which learning and reversal learning occurred. In the simple food seeking experiment, an agent representation of the model had to learn to find food rewards embedded in a green landmark in an environment. A yellow landmark which did not contain a reward was also included in the environment. On average, the agents approach to the green landmark increased over the duration of the run. In comparison, the agents approach to the yellow landmark decreased. The model's performance in the behavioural experiment, subject to simulated core and shell lesions were associated with and compared against real live experiments conducted by Parkinson et al. (2000). The model demonstrated similar behaviors to shell and core lesioned experiments conducted in vivo. In addition, the model's performance in serial reversal learning experiments were compared against empirical results

produced by Bushnell and Stanton (1991) and Watson et al. (2006).

Chapter 5 introduced the model as a modified actor-critic architecture. DA was employed in the model to acquire stimulus-action associations which resulted in reward delivery. When the reward was omitted, a feed-forward value switch was used to adjust actions. This feed-forward pathway represents the pathway between the shell and the cortical projections to the core via the mediodorsal nucleus of the thalamus and plays an essential role in facilitation and inhibition when rewards are omitted so that learned stimulus-response associations are not destroyed. A comparison was made between three different versions of the computational model categorised according to the rate by which unlearning occurred in the actor i.e. the rate of LTD in the core, and whether or not there was a feed-forward connection between the critic (shell) and actor (core). The model, with the MD feed-forward loop, performed comparatively better than the other two model versions. The results support the theory that rapid unlearning might not be the optimal option for demonstrating action selection. These observations suggest that classical actor-critic models are not sufficient models for behavioral flexibility conducted by the limbic system.

A variety of actor-critic models were described and compared against the current model. The roles of dopamine, the nucleus accumbens and the MD in behavioural flexibility were discussed. This work concludes by presenting some suggestions for future works.

6.8 Future Work

Some suggestions for the future, from integrating limbic structures into the model, to extending the model by connecting it with the dorsal striatum have been discussed.

6.8.1 The Limbic and Cortical Afferents

While the current work has focused on the cortico-striatal connectivity, the NAc is also innervated by the hippocampus (HPC) and the amygdala. The hippocampus has been implicated in the storage and the recall of new information based on configural learning in both space and time (Clark and Boutros, 1999). The amygdala comprises nuclei which are involved in emotional learning and expression. It has been associated especially with the acquisition and expression of conditioned fear and anxiety (Clark and Boutros, 1999; Cardinal et al., 2002a). It is involved in the Pavlovian conditioning of emotional responses. In particular, the basolateral nucleus of the amygdala (BLA) has been implicated in secondary conditioning, while the central nucleus of the amygdala has been implicated in conditioned orienting (Cardinal et al., 2002a). According to Cardinal et al. (2002a), the BLA is required for a conditioned stimulus (CS) to obtain information about the affective or motivational value of the unconditioned stimulus (US). The information obtained is then used to control different responses associated with fear or instrumental choice behaviour. The information obtained by the BLA can be relayed to the prefrontal cortex (PFC) as well as the NAc. The BLA has been implicated in facilitating cue-evoked reward seeking behaviour (Ishikawa et al., 2008). These limbic units could be integrated onto the NAc however, the mechanism by which the NAc selects and processes information from the limbic and cortical structures need to be addressed.

Anatomical and electrophysiological studies have shown converging afferents from both limbic and cortical afferents onto single NAc neurons (French and Totterdell, 2002). NAc neurons have also been observed to exist in two activity states. A hyperpolarised up state and a depolarised down state. Hippocampal activities are capable of driving these neurons into the up state. In addition, the HPC synchronises with the membrane states of the NAc (Goto and O'Donnell, 2001). French and Totterdell (2002) and Goto and O'Donnell (2001) suggest that it functions as a gate for the limbic and cortical afferents. The PFC inputs on the NAc show comparatively weak synchronizations

with the membrane state (Goto and O'Donnell, 2001). The HPC as well as the status of the NAc membrane potential could be implemented into an extended version of the model, to make selections between the limbic and cortical afferents.

The DA modulation of limbic and cortical inputs on the NAc are involved in goal directed behaviour. While the current model has focused on the cortical innervations to the NAc, it has been shown that tonic and phasic DA activity modulates hippocampal and PFC inputs via the respective activation of D1 and D2 receptors (Goto and Grace, 2005b). In particular, phasic DA selectively facilitates hippocampal inputs via the activation of D1 receptors. On the other hand, increasing and decreasing tonic activity attenuates and facilitates PFC inputs via the D2 receptor respectively. These findings suggest that an additional role of DA in the NAc is to selectively modulate limbic and cortical inputs. Goto and Grace (2005b) suggest that D2 receptor modulation of PFC inputs plays a role in the set shifting of response strategies.

The NAc neurons also receive converging inputs from the hippocampus and the amygdala. Tetanic stimulation of the hippocampal formation potentiated hippocampal evoked activity while suppressing the amygdala inputs. In addition, D1 receptor activation potentiates hippocampal evoked activity while D1 receptor inactivation blocked the suppression of amygdala evoked activity (Floresco et al., 2001).

While the current model emphasises the role of the MD in facilitating behavioural flexibility, the mechanism by which DA activity modulates cortical inputs via the D2 receptors could be employed to further improve the model. Cortico-limbic inputs on the NAc have been implicated in behaviours ranging from behavioural flexibility, to attentional and spatial learning, to secondary conditioning. The role and mechanisms by which each limbic and cortical regions influence on the NAc shell and core have not been established in the model. In addition, the processes by which the NAc sub-units select, obtain and transfer information from each afferent region could be analysed and im-

plemented into the model. These limbic afferents can be used to assign value to the stimuli that are processed by the cortical structures. For example, the hippocampus adds place information, while the amygdala '*tags*' emotional contexts such as fear to stimuli. The model can use such information to associate place fields with stimuli or to mimic avoidance behaviour.

6.8.2 The Dorsal Striatum and Basal Ganglia

In addition to the experimental observations obtained, a few assumptions were made which were used to develop the model surrounding the NAc circuitry. The core has been modelled to function similar to the dorsal striatum. The dorsal striatum as part of the basal ganglia, mediates actions by releasing inhibition. The core is assumed to achieve this by inhibiting the dorsolateral region of the ventral pallidum and as such dis-inhibiting the inhibitory influences on action. In this work, the release in behaviour by the core, (which is assumed to function similarly to the dorsal striatum,) has been simplified so that the core enables behaviour.

One major limitation of the model is that it has been developed at a systems level and tested in isolation from the dorsal striatum and limbic structures which contribute to an improved functionality of the model. This means that the model might fail in the lesion experiments conducted in chapter 4, if instead of partial lesions, complete lesions were used.

The relationship between the ventral and dorsal striatum is one such that the ventral striatum exerts control on procedures mediated by the dorsal striatal regions (Belin and Everitt, 2008). In studies involving drug addiction, Everitt et al. (2001) indicate how drug seeking behaviour progressively transfers from a goal-directed dimension, to a stimulus-response habit. This shift reflects a transition from the ventral to the dorsal striatal control over drug seeking (Belin and Everitt, 2008). It also indicates that a connection exists between ventral and dorsal regions. One pathway that connects the ventral to the dorsal striatum is the "*spiralling*" striato-nigral-striatal circuitry (Haber et al.,

2000; Belin and Everitt, 2008). This circuit includes the NAc shell which projects to the DA neurons of the VTA. The VTA neurons project to both the NAc shell and core. The NAc core also projects to DA neurons which in turn innervate the dorsal striatum (Haber et al., 2000). Belin and Everitt (2008) observed that by disabling the serial interactions between the ventral and dorsal striatal domains, drug seeking in rats trained to respond for drug under a second-order schedule of reinforcement was greatly and selectively decreased. These studies showed the importance of interactions between the core and the dopaminergic innervation of the dorsal striatum, in controlling established instrumental drug-seeking responses.

In addition to extending the model to integrate limbic afferents to the NAc, the model could be integrated into a systems level model of the basal ganglia so that the combined system's functionality is extended to mediate behaviours associated with the limbic and cortical regions. One example which shows how a limbic region transfers information to the dorsal striatum is described as follows: The interaction between the BLA (Whitelaw et al., 1996) and the core (Ito et al., 2004) is important for the acquisition of drug seeking through the regulation of conditioned reinforcement (Belin and Everitt, 2008). These mechanisms are dependent on the dorsolateral striatal DA which exert control over behaviour (Ito et al., 2002; Belin and Everitt, 2008). By studying how limbic nuclei such as the BLA interact with the NAc, (which in turn exerts control on the dorsal striatum,) the NAc shell and core circuitry and interaction could be improved and further developed as the limbic-motor interface that Mogenson et al. (1980) described.

6.8.3 The Sensitivity of the Model

A variety of parameters were used in the different simulations which could have been standardised based on more specific details such as the connectivity between each nuclei. The model could be defined according to set parameters which could then be implemented for all behavioral experiments. These set parameters may be generated by subjecting the model to different extremes

and analysing the performance of the model including its actor-critic versions presented in chapter 5 accordingly.

6.9 Conclusion

A model has been presented in this thesis that is based on the limbic circuitry. It supports a central role of the NAc in behavioral flexibility. Consequently, learning and reversal learning can be simulated by the model. The model suggests that the shell and core of the NAc have distinct properties. The shell plays an essential role in switching between basic behaviours by inhibiting irrelevant responses. The core on the other hand, enables behaviour in response to the reward predicting stimuli.

In this work, it has been proposed that dopamine acts distinctly on the shell and core. It enables "*learning*" and "*unlearning*" in these two sub-regions of the NAc. However, unlearning is assumed to occur at a significantly lower rate in the core than it does in the shell. In this way, the shell learns and unlearns rather quickly. This is a useful property for mediating flexibility. The shell influences the core through a pathway that corresponds to the mediodorsal nucleus of the thalamus. This in addition to the processes by which the shell and core learn and adapt are the main, novel and biologically inspired features which have not been implemented in previous computational models.

The model's ability to acquire associations and demonstrate behavioural flexibility has been shown in both open- and closed-loop experiments. In the open-loop experiments, the model accounts for a variety of classical conditioning phenomena, including rapid reacquisition. The model performs successfully in acquiring associations in the closed-loop reward seeking tasks. In addition, it effectively mimics the detrimental effects of the ventral striatal shell and core circuitry in discriminatory approach behavior. This is achieved because the shell and core employ different learning rates, and because a value switch has been implemented. By integrating these novel features, the model

was also capable of performing in a similar way to rats in serial reversal learning experiments. Finally, a comparison was made between the model and a variety of actor-critic versions of the model. The results generated showed that the model outperformed the other methods in terms of how many errors were made and how many serial reversals were achieved over a fixed duration.

Appendix A

The Filters

The Low- and Highpass Filters are represented in the Laplace domain as:

$$\text{Lowpass: } H_{LP}(s) = \frac{K}{(s+p_1)+(s+p_2)} \tag{A.1}$$

$$\text{Highpass: } H_{HP}(s) = \frac{s^2}{(s+p_1)+(s+p_2)} \tag{A.2}$$

K is a konstant and the poles are defined:

$$p_1 = a + jb \tag{A.3}$$

$$p_2 = a - jb \tag{A.4}$$

where the real (a) and imaginery (b) parts are defined by:

$$a = \frac{\pi f}{q} \tag{A.5}$$

$$b = \sqrt{(2\pi f^2) - (\frac{\pi f}{q})^2} \tag{A.6}$$

f and q correspond to the oscillation frquencies and q-factor of the filter respectively.

Appendix B

The Model Equations

The Reward System: LH is represented by the filtered reward signal r(t):

$$LH(t) = r(t) * h_{LP}(t) \tag{B.1}$$

The LH innervates the VTA:

$$VTA(t) = \frac{1 + \kappa \cdot LH(t)}{1 + \upsilon \cdot mVP(t) + \eta \cdot Shell(t)} \tag{B.2}$$

Which produces the burst and tonic DA activity:

$$burst(t) = \begin{cases} 1 & \text{if } [\chi_{burst} \cdot VTA(t) * h_{HP}(t)] \geq \theta_{burst}, \\ 0 & \text{otherwise.} \end{cases} \tag{B.3}$$

$$tonic(t) = \Theta_{tonic}[\chi_{tonic} \cdot [VTA(t) * h_{LP}(t)]] \tag{B.4}$$

The Indirect Pathways:

$$mVP(t) = \begin{cases} VP_{min} & \text{if } \frac{1}{1+\zeta \cdot shell(t)} < VP_{min}, \\ \frac{1}{1+\zeta \cdot shell(t)} & \text{otherwise.} \end{cases} \tag{B.5}$$

$$MD(t) = \theta_{MD}(1 - mVP(t)) \tag{B.6}$$

The Input System: The reflex and n predictive inputs are filtered:

$$\text{US input: } u_0(t) = h_{LP}(t) * x_0(t) \tag{B.7}$$

$$\text{predictive inputs: } u_{pre-j}(t) = h_{LP}(t) * x_j(t) \tag{B.8}$$

$0 < j \leq n$.

The predictive inputs are capable of maintaining persistent activity:

$$u_j(t) = \begin{cases} 0 & \text{if } LH_{reset}, \\ PA_{max} & \text{for period } T_{PA}, \\ & \text{if } u_{pre-j}(t) \geq PA_{max} \\ u_{pre-j}(t) & \text{otherwise.} \end{cases} \tag{B.9}$$

They make up the PFC cortical inputs to the core and OFC inputs to the shell:

$$PFC_0(t) = u_0(t) \tag{B.10}$$

$$PFC_j(t) = u_j(t) + \theta_{MD} MD(t) \tag{B.11}$$

$$OFC_j(t) = u_j(t) \tag{B.12}$$

The NAc:

$$\begin{aligned} \textit{core-}j &= [PFC_{0-j} \cdot \rho_{0-j} + \sum_{j=1}^{n} PFC_j \cdot \rho_j] \\ &- \sum_{k \neq j}^{n} \lambda \cdot \textit{core-}k \end{aligned} \tag{B.13}$$

$$shell = LH \cdot \omega_0 + \sum_{j=1}^{n} OFC_j(t) \cdot \omega_j \tag{B.14}$$

The Weight Change in the NAc:

The Core Weight:

$$\begin{aligned} \rho_X(t) \leftarrow \rho_X(t) &+ \mu_{core}(X_{PA}(t) \cdot \textit{core-}X(t)' \cdot burst(t) \cdot (limit - \rho_x(t))) \\ &- \epsilon_{core}(u_{X\text{-}distal}(t) \cdot tonic(t)) \end{aligned} \tag{B.15}$$

The Shell Weight:

$$\begin{aligned} \omega_X(t) \leftarrow \omega_X(t) &+ \mu_{shell}(X_{CS}(t) \cdot shell(t)' \cdot burst(t) \cdot (limit - \omega_x(t))) \\ &- \epsilon_{shell}(u_{X\text{-}distal}(t) \cdot tonic(t)) \end{aligned} \tag{B.16}$$

Appendix C

The Simulation Parameters

The parameters used in the model in the open-loop experiments in chapter 3 and in the reacquisition runs in chapter 5 are provided in table C.1.

The open-loop simulator can be downloaded from:

http://isg.elec.gla.ac.uk/maria/NeuronalNetworkSimulator.tar.gz

The closed-loop simulator can be downloaded from:

http://isg.elec.gla.ac.uk/maria/simulator.tar.gz

Table C.1: Open-loop simulation parameters

Unit	**Parameters**
Shell	Adaptive unit: $\mu_{shell} = 0.5$; $\epsilon_{shell} = 0.01$ $\omega_0 = 5$;
Core	Adaptive unit: $\mu_{core} = 0.5$; $\epsilon_{shell} = 0.0005$ $\lambda = -1$;
LH	LP Filter: $f = 0.01$; $q = 0.51$; $K = 1$

Continued on Next Page...

Table C.1 – Continued

Unit	Parameters
PFC_i $(0 \leq i \leq n)$ OFC_j $(0 < j \leq n)$	LP Filter: $f = 0.01$; $q = 0.51$; $K = 1$ $LH_{reset} = 0.1$ $T_{PA} = 0$ $PA_{max} = PFC_i$ $T_{PA} = 0$ $PA_{max} = OFC_j$
VTA	$\kappa = 0.8$; $\upsilon = 1$; $\eta = 1$
burst	HP Filter: $f = 0.1$; $q = 0.9$; $K = (2\pi f)^2$ $\chi_{burst} = 1$; $\theta_{burst} = 0.05$
tonic	LP Filter: $f = 0.01$; $q = 0.51$ $\chi_{tonic} = 0.1$; $\theta_{tonic} = 0$
mVP	$\zeta = 1$; $VP_{min} = 0.7$
MD	$\theta_{MD} = 0$

The parameters used in the model in the closed-loop experiments in chapter 4 are provided in table C.2. The parameters used in the model in the closed-loop comparison experiments in chapter 5 are provided in table C.3.

Table C.2: Closed-loop simulation parameters: The reward seeking & autoshaping experiments

Unit	**Full Model**	**Shell Lesion**	**Core Lesion**
Shell	Adaptive unit: $\mu_{shell} = 0.01$;$\epsilon_{shell} = 3e^{-4}$; $\omega_0 = 2$		
Core	Adaptive unit: $\mu_{core} = 1e^{-3}$; $\epsilon_{core} = 2e^{-5}$ $\lambda = -0.95$		
LH	LP Filter: $f = 0.01$; $q = 0.51$; $K = 1$		
PFC_i $(0 \leq i \leq n)$	$LH_{reset} = 0.01$; $T_{PA} = 0$; $PA_{max} = 0.5$		
OFC_j $(0 < j \leq n)$	$LH_{reset} = 0.01$; $T_{PA} = 1000$; $PA_{max} = 0.5$		
VTA	$\kappa = 1$; $\upsilon = 1$; $\eta = 0.5$	$\eta = 1e^{-4}$	$\eta = 0.5$
burst	HP Filter: $f = 0.01$; $q = 0.71$; $K = (2\pi f)^2$; $\chi_{burst} = 1$; $\theta_{burst} = 0.05$		
tonic	LP Filter: $f = 5e^{-4}$; $q = 0.51$;$K = 1$ $\chi_{tonic} = 1e^{-4}$; $\theta_{tonic} = 0$		
mVP	$\zeta = 1$; $VP_{min} = 0.7$		
MD	$\theta_{MD} = 10$		
Shell Efferent Shell-VP	5	0.2	5
Core Efferent Core-Enable motor	1	1	0.02

Table C.3: Closed-loop simulation parameters: The actor-critic comparison experiments

Unit	**pLTD-MD**	**pLTD**	**fLTD**
Shell	Adaptive unit: $\mu_{shell} = 0.01$;$\epsilon_{shell} = 3e^{-4}$; $\omega_0 = 2$		
Core	Adaptive unit: $\mu_{core} = 1e^{-3}$; $\lambda = -0.95$		
	$\epsilon_{core} = 1e^{-5}$	$\epsilon_{core} = 1e^{-5}$	$\epsilon_{core} = 1e^{-4}$
LH	LP Filter: $f = 0.01$; $q = 0.51$; $K = 1$		
PFC_i	$LH_{reset} = 0.01$; $T_{PA} = 0$; $PA_{max} = 0.5$		
$(0 \leq i \leq n)$			
OFC_j	$LH_{reset} = 0.01$; $T_{PA} = 1000$; $PA_{max} = 0.5$		
$(0 < j \leq n)$			
VTA	$\kappa = 1$; $\upsilon = 1$;		
	$\eta = 0.5$	$\eta = 1e^{-4}$	$\eta = 0.5$
burst	HP Filter: $f = 0.01$; $q = 0.71$; $K = (2\pi f)^2$; $\chi_{burst} = 1$; $\theta_{burst} = 0.05$		
tonic	LP Filter: $f = 5e^{-4}$; $q = 0.51$;$K = 1$ $\chi_{tonic} = 1e^{-4}$; $\theta_{tonic} = 0$		
mVP	$\zeta = 1$; $VP_{min} = 0.7$		
MD	$\theta_{MD} = 10$		
MD-PFC	1	0.65	0.65
Shell Efferent			
Shell-VP	5	0.2	5
Core Efferent			
Core-Enable motor	1	1	0.02

Bibliography

Abbott, A. Neuroscience: the molecular wake-up call. *Nature*, 447(7143):368–370.

Alcaro, A., Huber, R., and Panksepp, J. Behavioral functions of the mesolimbic dopaminergic system: an affective neuroethological perspective. *Brain Res Rev*, 56(2):283–321.

Alexander, G. E. and Crutcher, M. D. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci*, 13(7):266–271.

Alheid, G. F. and Heimer, L. New perspectives in basal forebrain organization of special relevance for neuropsychiatric disorders: the striatopallidal, amygdaloid, and corticopetal components of substantia innominata. *Neuroscience*, 27(1):1–39.

Annett, L. E., McGregor, A., and Robbins, T. W. The effects of ibotenic acid lesions of the nucleus accumbens on spatial learning and extinction in the rat. *Behav Brain Res*, 31(3):231–242.

Balkenius, C. and Morén, J. Computational models of classical conditioning: a comparative study. In Mayer, J.-A., R. H. L. W. S. W. and Blumberg, B., editors, *From Animals to Animats 5*. MIT Press.

Balleine, B. and Killcross, S. Effects of ibotenic acid lesions of the nucleus accumbens on instrumental action. *Behav Brain Res*, 65(2):181–193.

Bandura, A. Self-efficacy: Toward a unifying theory of behaviour change. *Phychological Review*, 84:191–215.

Barrot, M., Marinelli, M., Abrous, D. N., Rougé-Pont, F., Le Moal, M., and Piazza, P. V. The dopaminergic hyper-responsiveness of the shell of the nucleus accumbens is hormone-dependent. *Eur J Neurosci*, 12(3):973–979.

Barto, A. G. Adaptive critics and the basal ganglia. In *Models of Information Processing in the Basal Ganglia*, pages 215–232, Cambridge. MA:MIT Press.

Barto, A. G., Sutton, R. S., and Anderson, C. W. Neuronlike elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, 13:835–846.

Barto, A. G., Sutton, R. S., and Watkins, C. J. C. H. Learning and sequential decision making. pages 539–602.

Bassareo, V., De Luca, M. A., and Di Chiara, G. Differential impact of pavlovian drug conditioned stimuli on in vivo dopamine transmission in the rat accumbens shell and core and in the prefrontal cortex. *Psychopharmacology (Berl)*, 191(3):689–703.

Bassareo, V. and Di Chiara, G. Differential responsiveness of dopamine transmission to food-stimuli in nucleus accumbens shell/core compartments. *Neuroscience*, 89(3):637–641.

Belin, D. and Everitt, B. J. Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. *Neuron*, 57(3):432–441.

Berns, G. S. and Sejnowski, T. J. A computational model of how the basal ganglia produce sequences. *J Cogn Neurosci*, 10(1):108–121.

Birrell, J. M. and Brown, V. J. Medial frontal cortex mediates perceptual attentional set shifting in the rat. *J Neurosci*, 20(11):4320–4324.

Blaiss, C. A. and Janak, P. H. The nucleus accumbens core and shell are critical for the expression, but not the consolidation, of pavlovian conditioned approach. *Behav Brain Res*.

Boeijinga, P. H., Mulder, A. B., Pennartz, C. M., Manshanden, I., and Lopes da Silva, F. H. Responses of the nucleus accumbens following fornix/fimbria stimulation in the rat. identification and long-term potentiation of mono- and polysynaptic pathways. *Neuroscience*, 53(4):1049–1058.

Bouton, M. E. Differential control by context in the inflation and reinstatement paradigms. *Journal of Experimental Psychology: Animal Behavior Processes*, 10(1):56–74.

Bouton, M. E. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol Psychiatry*, 52(10):976–986.

Braitenberg, V. *Vehicles: Explorations In Synthetic Psychology*. MA:MIT Press, Cambridge.

Brog, J. S., Salyapongse, A., Deutch, A. Y., and Zahm, D. S. The patterns of afferent innervation of the core and shell in the "accumbens" part of the rat ventral striatum: immunohistochemical detection of retrogradely transported fluoro-gold. *J Comp Neurol*, 338(2):255–278.

Brown, J., Bullock, D., and Grossberg, S. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J Neurosci*, 19(23):10502–10511.

Brown, P. L. and Jenkins, H. M. Auto-shaping of the pigeon's key-peck. *J Exp Anal Behav*, 11(1):1–8.

Bushnell, P. J. and Stanton, M. E. Serial spatial reversal learning in rats: comparison of instrumental and automaintenance procedures. *Physiol Behav*, 50(6):1145–1151.

Calabresi, P., Centonze, D., Gubellini, P., Marfia, G. A., Pisani, A., Sancesario, G., and Bernardi, G. Synaptic transmission in the striatum: from plasticity to neurodegeneration. *Prog Neurobiol*, 61(3):231–265.

Calabresi, P., Maj, R., Mercuri, N. B., and Bernardi, G. Coactivation of d1 and d2 dopamine receptors is required for long-term synaptic depression in the striatum. *Neurosci Lett*, 142(1):95–99.

Calabresi, P., Maj, R., Pisani, A., Mercuri, N. B., and Bernardi, G. Long-term synaptic depression in the striatum: physiological and pharmacological characterization. *J Neurosci*, 12(11):4224–4233.

Calabresi, P., Mercuri, N., Stanzione, P., Stefani, A., and Bernardi, G. Intracellular studies on the dopamine-induced firing inhibition of neostriatal neurons in vitro: evidence for d1 receptor involvement. *Neuroscience*, 20(3):757–771.

Calabresi, P., Picconi, B., Tozzi, A., and Di Filippo, M. Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends Neurosci*, 30(5):211–219.

Calabresi, P., Pisani, A., Centonze, D., and Bernardi, G. Synaptic plasticity and physiological interactions between dopamine and glutamate in the striatum. *Neurosci Biobehav Rev*, 21(4):519–523.

Calabresi, P., Pisani, A., Mercuri, N. B., and Bernardi, G. Long-term potentiation in the striatum is unmasked by removing the voltage-dependent magnesium block of nmda receptor channels. *Eur J Neurosci*, 4(10):929–935.

Calabresi, P., Pisani, A., Mercuri, N. B., and Bernardi, G. The corticostriatal projection: from synaptic plasticity to dysfunctions of the basal ganglia. *Trends Neurosci*, 19(1):19–24.

Calaminus, C. and Hauber, W. Intact discrimination reversal learning but slowed responding to reward-predictive cues after dopamine d1 and d2 receptor blockade in the nucleus accumbens of rats. *Psychopharmacology (Berl)*, 191(3):551–566.

Calaminus, C. and Hauber, W. Guidance of instrumental behavior under reversal conditions requires dopamine d1 and d2 receptor activation in the orbitofrontal cortex. *Neuroscience*, 154(4):1195–1204.

Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev*, 26(3):321–352.

Cardinal, R. N., Parkinson, J. A., Lachenal, G., Halkerston, K. M., Rudarakanchana, N., Hall, J., Morrison, C. H., Howes, S. R., Robbins, T. W., and Everitt, B. J. Effects of selective excitotoxic lesions of the nucleus accumbens core, anterior cingulate cortex, and central nucleus of the amygdala on autoshaping performance in rats. *Behav Neurosci*, 116(4):553–567.

Carlsson, A. and Waldeck, B. A fluorimetric method for the determination of dopamine (3-hydroxytyramine). *Acta Physiol Scand*, 44(3-4):293–298.

Centonze, D., Gubellini, P., Picconi, B., Calabresi, P., Giacomini, P., and Bernardi, G. Unilateral dopamine denervation blocks corticostriatal ltp. *J Neurophysiol*, 82(6):3575–3579.

Cepeda, C., Buchwald, N. A., and Levine, M. S. Neuromodulatory actions of dopamine in the neostriatum are dependent upon the excitatory amino acid receptor subtypes activated. *Proc Natl Acad Sci U S A*, 90(20):9576–9580.

Cepeda, N. J., Cave, K. R., Bichot, N. P., and Kim, M. S. Spatial selection via feature-driven inhibition of distractor locations. *Percept Psychophys*, 60(5):727–746.

Charpier, S. and Deniau, J. M. In vivo activity-dependent plasticity at cortico-striatal connections: evidence for physiological long-term potentiation. *Proc Natl Acad Sci U S A*, 94(13):7036–7040.

Chaudhri, N., Sahuque, L. L., Cone, J. J., and Janak, P. H. Reinstated ethanol-seeking in rats is modulated by environmental context and requires the nucleus accumbens core. *Eur J Neurosci*, 28(11):2288–2298.

Chudasama, Y., Bussey, T. J., and Muir, J. L. Effects of selective thalamic and prelimbic cortex lesions on two types of visual discrimination and reversal learning. *Eur J Neurosci*, 14(6):1009–1020.

Churchill, L., Zahm, D. S., Duffy, P., and Kalivas, P. W. The mediodorsal nucleus of the thalamus in rats–ii. behavioral and neurochemical effects of gaba agonists. *Neuroscience*, 70(1):103–112.

Clark, D. L. and Boutros, N. N. *The brain and behavior : an introduction to behavioral neuroanatomy*. Blackwell Science, Massachusetts.

Contreras-Vidal, J. L. and Schultz, W. A predictive reinforcement model of dopamine neurons for learning approach behavior. *J Comput Neurosci*, 6(3):191–214.

Cools, R., Lewis, S. J., Clark, L., Barker, R. A., and Robbins, T. W. L-dopa disrupts activity in the nucleus accumbens during reversal learning in parkinson's disease. *Neuropsychopharmacology*, 32(1):180–189.

Corbit, L. H., Muir, J. L., and Balleine, B. W. The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. *J Neurosci*, 21(9):3251–3260.

Cragg, S. J. Meaningful silences: how dopamine listens to the ach pause. *Trends Neurosci*, 29(3):125–131.

Creese, I., Sibley, D. R., and Leff, S. Classification of dopamine receptors. *Adv Biochem Psychopharmacol*, 37:255–266.

Croft, A., Davison, R., and Hargreaves, M. *Engineering Mathematics A foundation for Electronic, Electrical, Communications and Systems Engineering*. Pearson Education Limited, Essex.

Dalia, A., Uretsky, N. J., and Wallace, L. J. Dopaminergic agonists administered into the nucleus accumbens: effects on extracellular glutamate and on locomotor activity. *Brain Res*, 788(1-2):111–117.

Dalley, J. W., Lääne, K., Theobald, D. E., Armstrong, H. C., Corlett, P. R., Chudasama, Y., and Robbins, T. W. Time-limited modulation of appetitive pavlovian memory by d1 and nmda receptors in the nucleus accumbens. *Proc Natl Acad Sci U S A*, 102(17):6189–6194.

Daw, N. D., Kakade, S., and Dayan, P. Opponent interactions between serotonin and dopamine. *Neural Netw*, 15(4-6):603–616.

Day, J. J. and Carelli, R. M. The nucleus accumbens and pavlovian reward learning. *Neuroscientist*, 13(2):148–159.

Dayan, P. Motivated reinforcement learning. *Advances in Neural Information Processing Systems*, 13.

Dayan, P. and Balleine, B. W. Reward, motivation, and reinforcement learning. *Neuron*, 36(2):285–298.

Deutch, A. Y. The regulation of subcortical dopamine systems by the prefrontal cortex: interactions of central dopamine systems and the pathogenesis of schizophrenia. *J Neural Transm Suppl*, 36:61–89.

Di Ciano, P., Cardinal, R. N., Cowell, R. A., Little, S. J., and Everitt, B. J. Differential involvement of nmda, ampa/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. *J Neurosci*, 21(23):9471–9477.

Di Filippo, M., Picconi, B., Tantucci, M., Ghiglieri, V., Bagetta, V., Sgobio, C., Tozzi, A., Parnetti, L., and Calabresi, P. Short-term and long-term plasticity at corticostriatal synapses: implications for learning and memory. *Behav Brain Res*, 199(1):108–118.

Dickinson, A., Smith, J., and Mirenowicz, J. Dissociation of pavlovian and instrumental incentive learning under dopamine antagonists. *Behav Neurosci*, 114(3):468–483.

Divac, I., Fonnum, F., and Storm-Mathisen, J. High affinity uptake of glutamate in terminals of corticostriatal axons. *Nature*, 266(5600):377–378.

Dorf, R. C. and Bishop, R. H. *Modern Control Systems*. Pearson Prentice Hall.

Durstewitz, D. and Seamans, J. K. The computational role of dopamine d1 receptors in working memory. *Neural Netw*, 15(4-6):561–572.

Egerton, A., Brett, R. R., and Pratt, J. A. Acute delta9-tetrahydrocannabinol-induced deficits in reversal learning: neural correlates of affective inflexibility. *Neuropsychopharmacology*, 30(10):1895–1905.

Everitt, B. J., Dickinson, A., and Robbins, T. W. The neuropsychological basis of addictive behaviour. *Brain Res Brain Res Rev*, 36(2-3):129–138.

Fiorillo, C. D., Tobler, P. N., and Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614):1898–1902.

Floresco, S. B. Dopaminergic regulation of limbic-striatal interplay. *J Psychiatry Neurosci*, 32(6):400–411.

Floresco, S. B., Blaha, C. D., Yang, C. R., and Phillips, A. G. Modulation of hippocampal and amygdalar-evoked activity of nucleus accumbens neurons by dopamine: cellular mechanisms of input selection. *J Neurosci*, 21(8):2851–2860.

Floresco, S. B., Braaksma, D. N., and Phillips, A. G. Thalamic-cortical-striatal circuitry subserves working memory during delayed responding on a radial arm maze. *J Neurosci*, 19(24):11061–11071.

Floresco, S. B., Ghods-Sharifi, S., Vexelman, C., and Magyar, O. Dissociable roles for the nucleus accumbens core and shell in regulating set shifting. *J Neurosci*, 26(9):2449–2457.

Floresco, S. B., McLaughlin, R. J., and Haluk, D. M. Opposing roles for the nucleus accumbens core and shell in cue-induced reinstatement of food-seeking behavior. *Neuroscience*, 154(3):877–884.

Floresco, S. B., West, A. R., Ash, B., Moore, H., and Grace, A. A. Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nat Neurosci*, 6(9):968–973.

Floresco, S. B., Zhang, Y., and Enomoto, T. Neural circuits subserving behavioral flexibility and their relevance to schizophrenia. *Behav Brain Res*.

French, S. J. and Totterdell, S. Hippocampal and prefrontal cortical inputs monosynaptically converge with individual projection neurons of the nucleus accumbens. *J Comp Neurol*, 446(2):151–165.

Freund, T. F., Powell, J. F., and Smith, A. D. Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. *Neuroscience*, 13(4):1189–1215.

Frey, P. W. and Ross, L. E. Classical conditioning of the rabbit eyelid response as a function of interstimulus interval. *J Comp Physiol Psychol*, 65(2):246–250.

Fuchs, R. A., Evans, K. A., Parker, M. C., and See, R. E. Differential involvement of the core and shell subregions of the nucleus accumbens in conditioned cue-induced reinstatement of cocaine seeking in rats. *Psychopharmacology (Berl)*, 176(3-4):459–465.

Fuchs, R. A., Ramirez, D. R., and Bell, G. H. Nucleus accumbens shell and core involvement in drug context-induced reinstatement of cocaine seeking in rats. *Psychopharmacology (Berl)*, 200(4):545–556.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol*, 61(2):331–349.

Galarraga, E., Hernández-López, S., Reyes, A., Barral, J., and Bargas, J. Dopamine facilitates striatal epsps through an l-type ca2+ conductance. *Neuroreport*, 8(9-10):2183–2186.

Garris, P. A., Ciolkowski, E. L., Pastore, P., and Wightman, R. M. Efflux of dopamine from the synaptic cleft in the nucleus accumbens of the rat brain. *J Neurosci*, 14(10):6084–6093.

Gerfen, C. R. Synaptic organization of the striatum. *J Electron Microsc Tech*, 10(3):265–281.

Ghitza, U. E., Fabbricatore, A. T., Prokopenko, V., Pawlak, A. P., and West, M. O. Persistent cue-evoked activity of accumbens neurons after prolonged abstinence from self-administered cocaine. *J Neurosci*, 23(19):7239–7245.

Gibbs, C. M., Latham, S. B., and Gormezano, I. Classical conditioning of the rabbit nictitating membrane response: effects of reinforcement schedule on response maintenance and resistance to extinction. *Anim Learn Behav*, 6(2):209–215.

Gillies, A. and Arbuthnott, G. Computational models of the basal ganglia. *Mov Disord*, 15(5):762–770.

Gonon, F. Prolonged and extrasynaptic excitatory action of dopamine mediated by d1 receptors in the rat striatum in vivo. *J Neurosci*, 17(15):5972–5978.

Gonon, F. and Sundstrom, L. Excitatory effects of dopamine released by impulse flow in the rat nucleus accumbens in vivo. *Neuroscience*, 75(1):13–18.

Goto, Y. and Grace, A. A. Dopamine-dependent interactions between limbic and prefrontal cortical plasticity in the nucleus accumbens: disruption by cocaine sensitization. *Neuron*, 47(2):255–266.

Goto, Y. and Grace, A. A. Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci*, 8(6):805–812.

Goto, Y. and Grace, A. A. Limbic and cortical information processing in the nucleus accumbens. *Trends Neurosci*, 31(11):552–558.

Goto, Y. and O'Donnell, P. Synchronous activity in the hippocampus and nucleus accumbens in vivo.

Goto, Y., Otani, S., and Grace, A. A. The yin and yang of dopamine release: a new perspective. *Neuropharmacology*, 53(5):583–587.

Grace, A. A. Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience*, 41(1):1–24.

Grace, A. A. The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction*, 95 Suppl 2:119–128.

Grace, A. A., Floresco, S. B., Goto, Y., and Lodge, D. J. Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends Neurosci*, 30(5):220–227.

Groenewegen, H. J. Organization of the afferent connections of the mediodorsal thalamic nucleus in the rat, related to the mediodorsal-prefrontal topography. *Neuroscience*, 24(2):379–431.

Groenewegen, H. J., Galis-de Graaf, Y., and Smeets, W. J. Integration and segregation of limbic cortico-striatal loops at the thalamic level: an experimental tracing study in rats. *J Chem Neuroanat*, 16(3):167–185.

Groenewegen, H. J., Vermeulen-Van der Zee, E., te Kortschot, A., and Witter, M. P. Organization of the projections from the subiculum to the ventral striatum in the rat. a study using anterograde transport of phaseolus vulgaris leucoagglutinin. *Neuroscience*, 23(1):103–120.

Gross, R. D. *Psychology: The Science of Mind and Behaviour*. Hodder & Stoughton, London.

Groves, P. M. A theory of the functional organization of the neostriatum and the neostriatal control of voluntary movement. *Brain Res*, 286(2):109–132.

Gubellini, P., Pisani, A., Centonze, D., Bernardi, G., and Calabresi, P. Metabotropic glutamate receptors and striatal synaptic plasticity: implications for neurological diseases. *Prog Neurobiol*, 74(5):271–300.

Gurney, K., Prescott, T. J., and Redgrave, P. A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biol Cybern*, 84(6):401–410.

Gurney, K., Prescott, T. J., and Redgrave, P. A computational model of action selection in the basal ganglia. ii. analysis and simulation of behaviour. *Biol Cybern*, 84(6):411–423.

Gurney, K. N., Humphries, M., Wood, R., Prescott, T. J., and Redgrave, P. Testing computational hypothesis of brain systems function: A case study with the basal ganglia. *Network: Computation in Neural Systems*, 15:263–290.

Haber, S. N., Fudge, J. L., and McFarland, N. R. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci*, 20(6):2369–2382.

Hall, J., Parkinson, J. A., Connor, T. M., Dickinson, A., and Everitt, B. J. Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating pavlovian influences on instrumental behaviour. *Eur J Neurosci*, 13(10):1984–1992.

Hebb, D. O. *The organization of behavior: A neurophychological study.* Wiley-Interscience, New York.

Horvitz, J. C. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96(4):651–656.

Houk, J. C., Adams, J. L., and Barto, A. G. A model of how the basal ganglia generates and uses neural signals that predict reinforcement. In Houk,

J. C., D. J. L. and Beiser, D. G., editors, *Models of Information Processing in the Basal Ganglia*, pages 249–274, Cambridge. MA:MIT Press.

Hsu, K. S., Huang, C. C., Yang, C. H., and Gean, P. W. Presynaptic d2 dopaminergic receptors mediate inhibition of excitatory synaptic transmission in rat neostriatum. *Brain Res*, 690(2):264–268.

Ikemoto, S. and Panksepp, J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Res Brain Res Rev*, 31(1):6–41.

Ishikawa, A., Ambroggi, F., Nicola, S. M., and Fields, H. L. Contributions of the amygdala and medial prefrontal cortex to incentive cue responding. *Neuroscience*, 155(3):573–584.

Ito, R., Dalley, J. W., Robbins, T. W., and Everitt, B. J. Dopamine release in the dorsal striatum during cocaine-seeking behavior under the control of a drug-associated cue. *J Neurosci*, 22(14):6247–6253.

Ito, R., Robbins, T. W., and Everitt, B. J. Differential control over cocaine-seeking behavior by nucleus accumbens core and shell. *Nat Neurosci*, 7(4):389–397.

Joel, D., Niv, Y., and Ruppin, E. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw*, 15(4-6):535–547.

Jones, D. L. and Mogenson, G. J. Nucleus accumbens to globus pallidus gaba projection: electrophysiological and iontophoretic investigations. *Brain Res*, 188(1):93–105.

Jones, M. W., Kilpatrick, I. C., and Phillipson, O. T. Regulation of dopamine function in the prefrontal cortex of the rat by the thalamic mediodorsal nucleus. *Brain Res Bull*, 19(1):9–17.

Jones, M. W., Kilpatrick, I. C., and Phillipson, O. T. Dopamine function in the prefrontal cortex of the rat is sensitive to a reduction of tonic gaba-

mediated inhibition in the thalamic mediodorsal nucleus. *Exp Brain Res*, 69(3):623–634.

Joseph, M. H., Peters, S. L., Moran, P. M., Grigoryan, G. A., Young, A. M., and Gray, J. A. Modulation of latent inhibition in the rat by altered dopamine transmission in the nucleus accumbens at the time of conditioning. *Neuroscience*, 101(4):921–930.

Kamin, L. Selective association and conditioning. pages 42–64.

Kandel, E., Schwartz, J., and Jessell, T. *Principles of Neural Science*. Prentice-Hall International Inc.

Kelley, A. E. Functional specificity of ventral striatal compartments in appetitive behaviors. *Ann N Y Acad Sci*, 877:71–90.

Kelley, A. E. Neural integrative activities of nucleus accumbens subregions in relation to learning and motivation. *Psychobiology*, 27:198–213.

Kelley, A. E. Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neurosci Biobehav Rev*, 27(8):765–776.

Kelley, A. E., Baldo, B. A., Pratt, W. E., and Will, M. J. Corticostriatal-hypothalamic circuitry and food motivation: integration of energy, action and reward. *Physiol Behav*, 86(5):773–795.

Kerr, J. N. and Wickens, J. R. Dopamine d-1/d-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. *J Neurophysiol*, 85(1):117–124.

Klopf, A. H., Weaver, S. E., and Morgan, J. S. A hierarchical network of control systems that learn: Modeling nervous system function during classical and instrumental conditioning. *Adaptive Behavior*, 1.

Kolodziejski, C., Porr, B., and Wörgötter, F. Mathematical properties of neuronal td-rules and differential hebbian learning: a comparison. *Biological Cybernetics*, 98.

Kötter, R. Postsynaptic integration of glutamatergic and dopaminergic signals in the striatum. *Prog Neurobiol*, 44(2):163–196.

Law-Tho, D., Desce, J. M., and Crepel, F. Dopamine favours the emergence of long-term depression versus long-term potentiation in slices of rat prefrontal cortex. *Neurosci Lett*, 188(2):125–128.

Lee, H. H., Choi, S. J., Cho, H. S., Kim, S. Y., and Sung, K. W. Dopamine modulates corticostriatal synaptic transmission through both d1 and d2 receptor subtypes in rat brain. *Korean J Physiol Pharmacol*, 9:263–268.

Levine, M. S., Li, Z., Cepeda, C., Cromwell, H. C., and Altemus, K. L. Neuromodulatory actions of dopamine on synaptically-evoked neostriatal responses in slices. *Synapse*, 24(1):65–78.

Lex, A. and Hauber, W. Dopamine d1 and d2 receptors in the nucleus accumbens core and shell mediate pavlovian-instrumental transfer. *Learn Mem*, 15(7):483–491.

Ljungberg, T., Apicella, P., and Schultz, W. Responses of monkey dopamine neurons during learning of behavioral reactions. *J Neurophysiol*, 67(1):145–163.

Maeno, H. Dopamine receptors in canine caudate nucleus. *Mol Cell Biochem*, 43(2):65–80.

Maldonado-Irizarry, C. S., Swanson, C. J., and Kelley, A. E. Glutamate receptors in the nucleus accumbens shell control feeding behavior via the lateral hypothalamus. *J Neurosci*, 15(10):6779–6788.

Malenka, R. C. and Bear, M. F. Ltp and ltd: an embarrassment of riches. *Neuron*, 44(1):5–21.

Matsuda, Y., Marzo, A., and Otani, S. The presence of background dopamine signal converts long-term synaptic depression to potentiation in rat prefrontal cortex. *J Neurosci*, 26(18):4803–4810.

McKittrick, C. R. and Abercrombie, E. D. Catecholamine mapping within nucleus accumbens: differences in basal and amphetamine-stimulated efflux of norepinephrine and dopamine in shell and core. *J Neurochem*, 100(5):1247–1256.

Means, L. W., Hershey, A. E., Waterhouse, G. J., and Lane, C. J. Effects of dorsomedial thalamic lesions on spatial discrimination reversal in the rat. *Physiol Behav*, 14(6):725–729.

Mercuri, N., Bernardi, G., Calabresi, P., Cotugno, A., Levi, G., and Stanzione, P. Dopamine decreases cell excitability in rat striatal neurons by pre- and postsynaptic mechanisms. *Brain Res*, 358(1-2):110–121.

Meredith, G. E. The synaptic framework for chemical signaling in nucleus accumbens. *Ann N Y Acad Sci*, 877:140–156.

Mogenson, G. J., Jones, D. L., and Yim, C. Y. From motivation to action: functional interface between the limbic system and the motor system. *Prog Neurobiol*, 14(2-3):69–97.

Mogenson, G. J., Yang, C. R., and Yim, C. Y. Influence of dopamine on limbic inputs to the nucleus accumbens. *Ann N Y Acad Sci*, 537:86–100.

Montague, P. R., Dayan, P., and Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive hebbian learning. *J Neurosci*, 16(5):1936–1947.

Moore, R. Y. and Bloom, F. E. Central catecholamine neuron systems: anatomy and physiology of the dopamine systems. *Annu Rev Neurosci*, 1:129–169.

Nambu, A. Basal ganglia: Physiological circuits. In *Encyclopedia of Neuroscience*, pages 111–117.

Napier, R. M., Macrae, M., and Kehoe, E. J. Rapid reaquisition in conditioning of the rabbit's nictitating membrane response. *J Exp Psychol Anim Behav Process*, 18(2):182–192.

Nicola, S. M. The nucleus accumbens as part of a basal ganglia action selection circuit. *Psychopharmacology (Berl)*, 191(3):521–550.

Nicola, S. M., Surmeier, J., and Malenka, R. C. Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annu Rev Neurosci*, 23:185–215.

Nicola, S. M., Yun, I. A., Wakabayashi, K. T., and Fields, H. L. Firing of nucleus accumbens neurons during the consummatory phase of a discriminative stimulus task depends on previous reward predictive cues. *J Neurophysiol*, 91(4):1866–1882.

Niv, Y., Daw, N. D., Joel, D., and Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl)*, 191(3):507–520.

O'Donnell, P. and Grace, A. A. Tonic d2-mediated attenuation of cortical excitation in nucleus accumbens neurons recorded in vitro. *Brain Res*, 634(1):105–112.

Olds, J. and Milner, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J Comp Physiol Psychol*, 47(6):419–427.

Ongür, D. and Price, J. L. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex*, 10(3):206–219.

O'Reilly, R. C., Frank, M. J., Hazy, T. E., and Watz, B. Pvlv: the primary value and learned value pavlovian learning algorithm. *Behav Neurosci*, 121(1):31–49.

Papp, M. and Bal, A. Separation of the motivational and motor consequences of 6-hydroxydopamine lesions of the mesolimbic or nigrostriatal system in rats. *Behav Brain Res*, 23(3):221–229.

Parent, A., Sato, F., Wu, Y., Gauthier, J., Lévesque, M., and Parent, M. Organization of the basal ganglia: the importance of axonal collateralization. *Trends Neurosci*, 23(10 Suppl):20–27.

Park, J. and Choi, J. S. Long-term synaptic changes in two input pathways into the lateral nucleus of the amygdala underlie fear extinction. *Learn Mem*, 17(1):23–34.

Parkinson, J. A., Dalley, J. W., Cardinal, R. N., Bamford, A., Fehnert, B., Lachenal, G., Rudarakanchana, N., Halkerston, K. M., Robbins, T. W., and Everitt, B. J. Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behav Brain Res*, 137(1-2):149–163.

Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W., and Everitt, B. J. Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by d-amphetamine. *J Neurosci*, 19(6):2401–2411.

Parkinson, J. A., Willoughby, P. J., Robbins, T. W., and Everitt, B. J. Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs pavlovian approach behavior: further evidence for limbic cortical-ventral striatopallidal systems. *Behav Neurosci*, 114(1):42–63.

Parsons, L. H. and Justice, J. B. Extracellular concentration and in vivo recovery of dopamine in the nucleus accumbens using microdialysis. *J Neurochem*, 58(1):212–218.

Passetti, F., Chudasama, Y., and Robbins, T. W. The frontal cortex of the rat and visual attentional performance: dissociable functions of distinct medial prefrontal subregions. *Cereb Cortex*, 12(12):1254–1268.

Patterson, M. M., Olah, J., and Clement, J. Classical nictitating membrane conditioning in the awake, normal, restrained cat. *Science*, 196(4294):1124–1126.

Pavlov, I, P. *Conditioned reflexes*. Oxford University Press, Oxford.

Pawlak, V. and Kerr, J. N. Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. *J Neurosci*, 28(10):2435–2446.

Pearce, J. M. *Animal learning and cognition, An introduction*. Psychology Press, Sussex.

Pennartz, C. M., Ameerun, R. F., Groenewegen, H. J., and Lopes da Silva, F. H. Synaptic plasticity in an in vitro slice preparation of the rat nucleus accumbens. *Eur J Neurosci*, 5(2):107–117.

Peters, J., LaLumiere, R. T., and Kalivas, P. W. Infralimbic prefrontal cortex is responsible for inhibiting cocaine seeking in extinguished rats. *J Neurosci*, 28(23):6046–6053.

Phillips, A. G., Brooke, S. M., and Fibiger, H. C. Effects of amphetamine isomers and neuroleptics on self-stimulation from the nucleus accumbens and dorsal noradrenergic bundle. *Brain Res*, 85(1):13–22.

Phillips, G. D., Robbins, T. W., and Everitt, B. J. Bilateral intra-accumbens self-administration of d-amphetamine: antagonism with intra-accumbens sch-23390 and sulpiride. *Psychopharmacology (Berl)*, 114(3):477–485.

Picconi, B., Pisani, A., Barone, I., Bonsi, P., Centonze, D., Bernardi, G., and Calabresi, P. Pathological synaptic plasticity in the striatum: implications for parkinson's disease. *Neurotoxicology*, 26(5):779–783.

Porr, B. *Sequence-Learning in a Self-Referential Closed-Loop Behavioural System*. PhD thesis.

Porr, B. and Wörgötter, F. Temporal hebbian learning in rate-coded neural networks: A theoretical approach towards classical conditioning. In Dorffner, G., Bischof, H., and Hornik, K., editors, *ICANN*, volume 2130 of *Lecture Notes in Computer Science*, pages 1115–1120. Springer.

Porr, B. and Wörgötter, F. Isotropic Sequence Order learning. *Neural Comp.*, 15:831–864.

Porr, B. and Wörgötter, F. Temporal sequence learning, prediction, and control - a review of different models and their relation to biological mechanisms. *Neural Computation*.

Porr, B. and Wörgötter, F. Learning with "relevance": Using a third factor to stabilise hebbian learning. *Neural Computation*.

Prescott, T. J., Gonzalez, F. M. M., Gurney, K., D., H. M., and Redgrave, P. A robot model of the basal ganglia: Behavior and intrinsic processing. *Neural Networks*, 19(1):31–61.

Rebec, G. V., Christensen, J. R., Guerra, C., and Bardo, M. T. Regional and temporal differences in real-time dopamine efflux in the nucleus accumbens during free-choice novelty. *Brain Res*, 776(1-2):61–67.

Redgrave, P. and Gurney, K. The short-latency dopamine signal: a role in discovering novel actions? *Nat Rev Neurosci*, 7(12):967–975.

Redgrave, P., Prescott, T. J., and Gurney, K. The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89(4):1009–1023.

Redgrave, P., Prescott, T. J., and Gurney, K. Is the short-latency dopamine response too short to signal reward error? *Trends Neurosci*, 22(4):146–151.

Rescorla, R. A. Experimental extinction. In Klein, S. B. and Mowrer, R. R., editors, *Handbook of Contemporary Learning Theories*, pages 119–154, Mahwah, NJ. Lawrence Erlbaum Associates.

Rescorla, R. A. and Heth, C. D. Reinstatement of fear to an extinguished conditioned stimulus. *J Exp Psychol Anim Behav Process*, 1(1):88–96.

Rescorla, R.A., W. A. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II*, pages 64–99.

Reynolds, J. N. and Wickens, J. R. Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw*, 15(4-6):507–521.

Reynolds, S. M. and Berridge, K. C. Fear and feeding in the nucleus accumbens shell: rostrocaudal segregation of gaba-elicited defensive behavior versus eating behavior. *J Neurosci*, 21(9):3261–3270.

Robbins, T. W., Cador, M., Taylor, J. R., and Everitt, B. J. Limbic-striatal interactions in reward-related processes. *Neurosci Biobehav Rev*, 13(2-3):155–162.

Robbins, T. W. and Everitt, B. J. Neurobehavioural mechanisms of reward and motivation. *Curr Opin Neurobiol*, 6(2):228–236.

Roberts, D. C., Corcoran, M. E., and Fibiger, H. C. On the role of ascending catecholaminergic systems in intravenous self-administration of cocaine. *Pharmacol Biochem Behav*, 6(6):615–620.

Roberts, P. Computational consequences of temporally asymmetric learning rules: I.differential hebbian learning. *J. Comput. Neurosci.*, 7:235–246.

Romo, R. and Schultz, W. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J Neurophysiol*, 63(3):592–606.

Salamone, J. D., Cousins, M. S., and Snyder, B. J. Behavioral functions of nucleus accumbens dopamine: empirical and conceptual problems with the anhedonia hypothesis. *Neurosci Biobehav Rev*, 21(3):341–359.

Schmajuk, N. Brain-behaviour relationships in latent inhibition: a computational model. *Neurosci Biobehav Rev*, 29(6):1001–1020.

Schmajuk, N., Cox, L., and Gray, J. Brain-behaviour relationships in latent inhibition: a computational model. *Behavioural Brain Research*, 118(2):123–141.

Schmajuk, N., Lam, Y., and Gray, J. Latent inhibition: a neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes*, 22(3):321–349.

Schneiderman, N. and Gormezano, I. Conditioning of the nictitating membrane of the rabbit as a function of cs-us interval. *J Comp Physiol Psychol*, 57:188–195.

Schotanus, S. M. and Chergui, K. Dopamine d1 receptors and group i metabotropic glutamate receptors contribute to the induction of long-term potentiation in the nucleus accumbens. *Neuropharmacology*, 54(5):837–844.

Schultz, W. The phasic reward signal of primate dopamine neurons. *Adv Pharmacol*, 42:686–690.

Schultz, W., Apicella, P., and Ljungberg, T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J Neurosci*, 13(3):900–913.

Schultz, W., Dayan, P., and Montague, P. R. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599.

Schultz, W. and Dickinson, A. Neuronal coding of prediction errors. *Annu Rev Neurosci*, 23:473–500.

Schultz, W., Tremblay, L., and Hollerman, J. R. Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereb Cortex*, 10(3):272–284.

Shepherd, M. S. *The synaptic organization of the brain*. Oxford University Press, New York; Oxford.

Silva, F. J., Silva, K. M., and Pear, J. J. Sign- versus goal-tracking: effects of conditioned-stimulus-to-unconditioned-stimulus distance. *J Exp Anal Behav*, 57(1):17–31.

Smith, A. D. and Bolam, J. P. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends Neurosci*, 13(7):259–265.

Smith, M. C., Coleman, S. R., and Gormezano, I. Classical conditioning of the rabbit's nictitating membrane response at backward, simultaneous, and forward cs-us intervals. *J Comp Physiol Psychol*, 69(2):226–231.

Solomon, P. R. and Staton, D. M. Differential effects of microinjections of d-amphetamine into the nucleus accumbens or the caudate putamen on the rat's ability to ignore an irrelevant stimulus. *Biol Psychiatry*, 17(6):743–756.

Steinfels, G. F., Heym, J., Strecker, R. E., and Jacobs, B. L. Behavioral correlates of dopaminergic unit activity in freely moving cats. *Brain Res*, 258(2):217–228.

Stern, C. E. and Passingham, R. E. The nucleus accumbens in monkeys (macaca fascicularis). iii. reversal learning. *Exp Brain Res*, 106(2):239–247.

Stratford, T. R. and Kelley, A. E. Gaba in the nucleus accumbens shell participates in the central regulation of feeding behavior. *J Neurosci*, 17(11):4434–4440.

Stratford, T. R. and Kelley, A. E. Evidence of a functional relationship between the nucleus accumbens shell and lateral hypothalamus subserving the control of feeding behavior. *J Neurosci*, 19(24):11040–11048.

Suri, R. E. and Schultz, W. Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Exp Brain Res*, 121(3):350–354.

Suri, R. E. and Schultz, W. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience*, 91(3):871–890.

Surmeier, D. J., Ding, J., Day, M., Wang, Z., and Shen, W. D1 and d2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends Neurosci*, 30(5):228–235.

Sutton, R. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44.

Sutton, R. and Barto, A. Simulation of anticipatory responses in classical conditioning by a neuron-like adaptive element. *Behavioural Brain Research*, 4(3):221–235.

Sutton, R. S. *Temporal credit assignment in reinforcement learning*. PhD thesis.

Sutton, R. S. and Barto, A. A temporal-difference model of classical conditioning. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, pages 355–378, Seattle, Washington.

Sutton, R. S. and Barto, A. Time-derivative models of Pavlovian reinforcement. In Gabriel, M. and Moore, J., editors, *Learning and Computational Neuroscience*, pages 497–537. MIT-press, Cambridge, MA.

Sutton, R. S. and Barto, A. G. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev*, 88(2):135–170.

Sutton, R. S. and Barto, A. G. *Reinforcement learning: an introduction*. MIT, Cambridge, MA.

Taghzouti, K., Louilot, A., Herman, J. P., Le Moal, M., and Simon, H. Alternation behavior, spatial discrimination, and reversal disturbances following 6-hydroxydopamine lesions in the nucleus accumbens of the rat. *Behav Neural Biol*, 44(3):354–363.

Taghzouti, K., Simon, H., Louilot, A., Herman, J. P., and Le Moal, M. Behavioral study after local injection of 6-hydroxydopamine into the nucleus accumbens in the rat. *Brain Res*, 344(1):9–20.

Thompson, A., Porr, B., and Wörgötter, F. Stabilising hebbian learning with a third factor in a food retrieval task. In Nolfi, S., editor, *Proceedings of the Ninth International Conference on Simulation of Adaptive Behavior, SAB, (LNAI 4095)*, pages 313–322. Springer.

Thompson, A., Porr, B., and Wörgötter, F. Learning and reversal in the sub-cortical limbic system: A computational model.

Thompson, A. M., Porr, B., Kolodziejski, C., and Wörgötter, F. Second order conditioning in the sub-cortical nuclei of the limbic system. In *SAB '08: Proceedings of the 10th international conference on Simulation of Adaptive Behavior*, pages 189–198, Berlin, Heidelberg. Springer-Verlag.

Thorndike, E. L. *Animal Intelligence*. Macmillan.

Tseng, K. Y., Snyder-Keller, A., and O'Donnell, P. Dopaminergic modulation of striatal plateau depolarizations in corticostriatal organotypic cocultures. *Psychopharmacology (Berl)*, 191(3):627–640.

Umemiya, M. and Raymond, L. A. Dopaminergic modulation of excitatory postsynaptic currents in rat neostriatal neurons. *J Neurophysiol*, 78(3):1248–1255.

Utter, A. A. and Basso, M. A. The basal ganglia: an overview of circuits and function. *Neurosci Biobehav Rev*, 32(3):333–342.

Verschure, P. F., Voegtlin, T., and Douglas, R. J. Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425(6958):620–624.

Wakabayashi, K. T., Fields, H. L., and Nicola, S. M. Dissociation of the role of nucleus accumbens dopamine in responding to reward-predictive cues and waiting for reward. *Behav Brain Res*, 154(1):19–30.

Wang, Z., Kai, L., Day, M., Ronesi, J., Yin, H. H., Ding, J., Tkatch, T., Lovinger, D. M., and Surmeier, D. J. Dopaminergic control of corticostriatal long-term synaptic depression in medium spiny neurons is mediated by cholinergic interneurons. *Neuron*, 50(3):443–452.

Watson, D. J., Sullivan, J. R., Frank, J. G., and Stanton, M. E. Serial reversal learning of position discrimination in developing rats. *Dev Psychobiol*, 48(1):79–94.

Weiner, I. The "two-headed" latent inhibition model of schizophrenia: modeling positive and negative symptoms and their treatment. *Psychopharmacology (Berl)*, 169(3-4):257–297.

Weiner, I., Izraeli-Telerant, A., and Feldon, J. Latent inhibition is not affected by acute or chronic administration of 6 mg/kg dl-amphetamine. *Psychopharmacology (Berl)*, 91(3):345–351.

Weiner, I. and Joel, D. I. Dopamine in schizophrenia: Dysfunctional information processing in basal gangliathalamocortical split circuits. In *Handbook of experimental pharmacology*, pages 417–472.

Welsh, J. P. and Harvey, J. A. Cerebellar lesions and the nictitating membrane reflex: performance deficits of the conditioned and unconditioned response. *J Neurosci*, 9(1):299–311.

Whitelaw, R. B., Markou, A., Robbins, T. W., and Everitt, B. J. Excitotoxic lesions of the basolateral amygdala impair the acquisition of cocaine-seeking behaviour under a second-order schedule of reinforcement. *Psychopharmacology (Berl)*, 127(3):213–224.

Widrow, B. and Hoff, M. Adaptive switching circuits. 4:96–104.

Wise, R. A. Drug-activation of brain reward pathways. *Drug Alcohol Depend*, 51(1-2):13–22.

Wise, R. A. Dopamine, learning and motivation. *Nat Rev Neurosci*, 5(6):483–494.

Wise, R. A. and Rompre, P. P. Brain dopamine and reward. *Annu Rev Psychol*, 40:191–225.

Wise, R. A., Spindler, J., deWit, H., and Gerberg, G. J. Neuroleptic-induced "anhedonia" in rats: pimozide blocks reward quality of food. *Science*, 201(4352):262–264.

Witten, I. H. An adaptive optimal controller for discrete-time markov environments. *Information and Control*, 34(3):483–494.

Yin, H. H., Davis, M. I., Ronesi, J. A., and Lovinger, D. M. The role of protein synthesis in striatal long-term depression. *J Neurosci*, 26(46):11811–11820.

Yin, H. H., Ostlund, S. B., and Balleine, B. W. Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur J Neurosci*, 28(8):1437–1448.

Yun, I. A., Nicola, S. M., and Fields, H. L. Contrasting effects of dopamine and glutamate receptor antagonist injection in the nucleus accumbens suggest a neural mechanism underlying cue-evoked goal-directed behavior. *Eur J Neurosci*, 20(1):249–263.

Zahm, D. S. An integrative neuroanatomical perspective on some subcortical substrates of adaptive responding with emphasis on the nucleus accumbens. *Neurosci Biobehav Rev*, 24(1):85–105.

Zahm, D. S. and Brog, J. S. On the significance of subterritories in the "accumbens" part of the rat ventral striatum. *Neuroscience*, 50(4):751–767.

Zahm, D. S. and Heimer, L. Two transpallidal pathways originating in the rat nucleus accumbens. *J Comp Neurol*, 302(3):437–446.

Zahm, D. S. and Heimer, L. Specificity in the efferent projections of the nucleus accumbens in the rat: comparison of the rostral pole projection patterns with those of the core and shell. *J Comp Neurol*, 327(2):220–232.